Historic, Archive Document

Do not assume content reflects current scientific knowledge, policies, or practices.

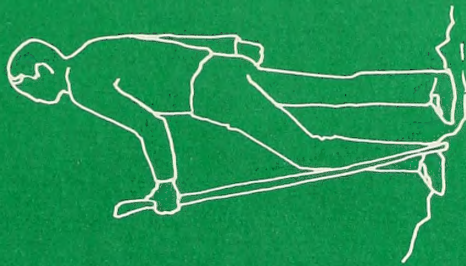# Measuring Landscape Esthetics: The Scenic Beauty Estimation Method

Terry C. Daniel
Ron S. Boster

Measuring Landscape Esthetics: The Scenic Beauty Estimation Method
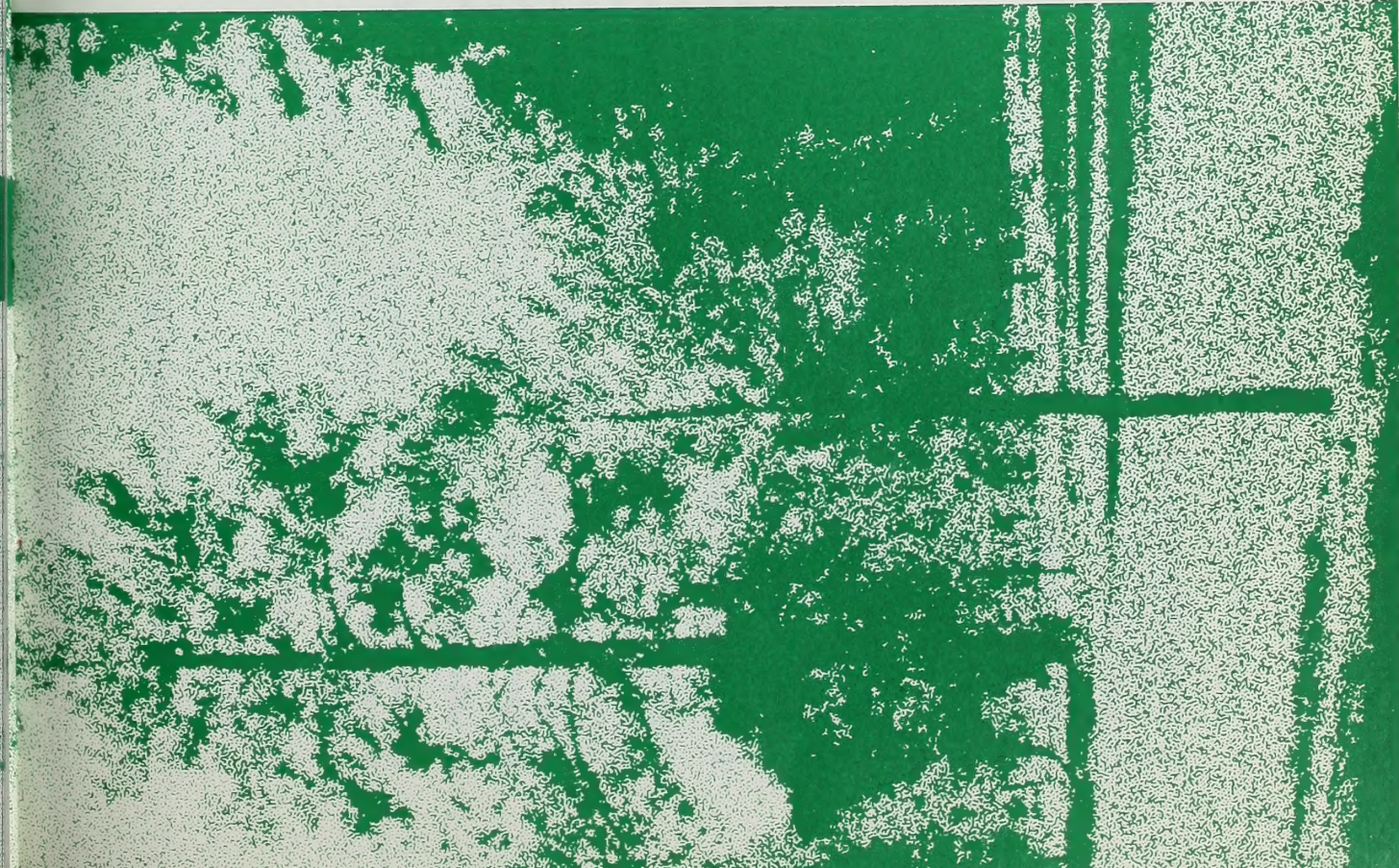
Terry C. Daniel
Associate Professor of Psychology
University of Arizona, Tucson

and

Ron S. Boster
Principal Economist
Rocky Mountain Forest and Range Experiment
Station, central headquarters maintained at Fort
Collins in cooperation with Colorado State
University

LANDSCAPE EVALUATION: OVERVIEW

Public sentiment and legislative mandate require that esthetic and other intangible consequences of public land use be considered. Landscape scenic beauty is one of the most important of our natural resources. Of the many resources we use, preserve, and try to improve, scenic beauty has proven one of the most difficult to measure in an objective, scientific manner. No doubt this is because beauty is only partially defined by characteristics of the environment, and depends, in large part, upon human judgement.

National Forests are a significant source of scenic beauty, and management must be responsive to the value of this resource. Assessment of scenic beauty and of management impacts on scenic beauty has, however, posed a difficult problem for public land managers. Meaningful indicators of public esthetic preferences are necessary to comprehensive, multi-use planning and management of our National Forests. The major purpose of this paper is to describe the development of a technique for measuring scenic beauty in terms of public perceptual judgment. In this context, our five specific objectives are:

1. To discuss the need for systematic, objective measurement and prediction of the scenic beauty of forest landscapes and related wildlands.

The esthetic qualities of forests and wildlands have long been recognized as important. Much of the recent upsurge in environmental concern has focused on intangibles such as scenic beauty. As is the case for most environmental intangibles, improved means of measurement and prediction are needed. A discussion of the need for new and improved methods for considering the scenic resource is an appropriate beginning to this report.

2. To review and comment on selected approaches to the problem of scenic beauty measurement and prediction.

Until recently, designers, planners, and decision-makers have relied to a large extent on either their own intuition or the intuition of others to assess scenic beauty. Predicting esthetic consequences of land management options has been equally subjective. Recently, a number of approaches and techniques have been offered for measuring and, to a

lesser extent, for predicting scenic beauty. A discussion of those approaches is needed to place the general problem in perspective.

3. To describe — for researchers, resource planners, and decision-makers — a methodology for measuring public esthetic preferences. The scenic Beauty Estimation Method — SBE for short — has evolved over a 4 year period of rigorous experimentation. We believe that a discussion of the theory underlying SBE is important to a more complete understanding of the method and the results presented.

4. To illustrate the SBE method and report the results of an application of the method with several user, interest, and professional groups. After development and initial testing, the SBE Method was used to determine similarities and differences in the perceptual preference of several public groups. These groups are all concerned with wildland management, some more directly than others.

5. To illustrate and suggest several extensions of the basic methodology which have relevant management applications. We have found that the basic SBE methodology is flexible, and may be used in several ways to aid designers, planners, and decision-makers. While these applications are still in the research phase, they are reported to stimulate thought and to obtain critical comments at an early stage of development.

These objectives are addressed in essentially independent sections of this report. Thus, the reader having a special interest in one particular objective can turn his attention to the appropriate section. In addition, we have separated discussions of the more technical aspects of our research from the main narrative. This material is supplementary, and is printed in green adjacent to appropriate narrative sections.

Why Measure Scenic Beauty?

Why indeed? This question, in one form or another, has come up frequently in our interaction with various public groups. Rarely does the question come from public land planners and decision-makers, who generally seem to agree upon the desirability of rating, measuring, or otherwise qualifying scenic beauty.

What's in a Name?

We have attempted to take an open-minded approach to naming the environmental property perceived. Several descriptive terms have been used. For example, Forest Service landscape architects seem to prefer the term "natural beauty", but have few qualms about substituting the terms "esthetics" or "landscape quality".

Our preference is "Scenic Beauty". We feel it is the most precise, although we frequently employ other terms, such as natural beauty, landscape esthetics, or scenic resource, to break the monotony. The word "natural" technically excludes many common components of wildland scenes that most people would agree "belong", such as rustic fences, corrals, and other human artifacts that are generally acceptable in a "natural" setting. Besides, few truly "natural" areas (untouched by human hands) exist. The term "esthetics" is often construed to involve more than just the visual senses (Shafer 1969). The term "value" has the cold connotation of money, while "quality" is somewhat nebulous (high quality, low quality, or a "quality", meaning property, of a landscape). Because "beauty" can be other than visual, we have chosen "scenic beauty".

While you and I have lips and voices
to kiss and sing with
Who cares if some (characterization deleted)
invents an instrument to measure spring with?

e. e. cummings

There are three main reasons for determining relative esthetic preferences for landscapes resulting from forest management practices:

- Better integration with other resources and products.
- Better justification for land use decisions.
- Restoration of the client-architect relationship.

Better Integration with Other Resources and Products

Planning and management of public lands is becoming increasingly complex. Perhaps the most vexing problem concerns how best to integrate the traditional economic concerns — typically represented by timber, forage, and water — with the less tangible "products", such as esthetics, wildlife, and recreation. While some argue that intangibles can never be properly quantified, many others, ourselves included, believe there is considerable opportunity for systematically incorporating intangibles into the planning process. Indeed, our studies make us optimistic that esthetic preferences can be quantified and predicted with no less accuracy than water, timber, and forage yields.

The scenic resource is but one of many. It should not be unnecessarily isolated or otherwise treated uniquely in the decision-making process. However, quantification need not require putting a dollar or other commensurate value on something as inherently non-pecuniary as scenic beauty. Rather, quantification should be considered as providing a useful information base for making more enlightened, informed decisions.

Better Justification for Land Use Decisions

Some judgments in natural resource decision-making are necessarily subjective and probably always will be. However, resources that can be objectively measured should be, so that decisions can be based on as much substantive evidence as possible. Because the public is seeking more information about and more involvement in public land planning, decision-makers are under increased pressure to better justify management decisions. Unfortunately, the reasons offered for land use decisions often take the form of rationalizations, even if well intended.

For the general public, the most immediate and direct impact of land management is visual. It is not surprising, then, that much of the reaction to public land management practices concerns esthetics (the clearcutting controversy, for instance). Partly in response to this public concern, numerous techniques purporting to inventory and/or measure scenic quality have been developed. The proliferation of such techniques (and to a lesser extent, their use by managers and planners) is a strong indication that better justification for land use decisions is needed.

There are many illustrations of the need for better justification of management actions. A timber sale provides an excellent example. With few exceptions, the immediate and short-term impacts on scenic beauty from a timber sale are negative relative to preharvest conditions. Most forest managers "know" that vegetative recovery can, in time, alleviate initially negative esthetic effects. Objective measurement, however, would provide better justification to an increasingly concerned and skeptical public than do intuitive assumptions or unsupported expert opinions.

Restoration of the Client-Architect Relationship

The number of landscape architects in public land managing agencies is increasing, as is their professional role. For example, the Forest Service employs more landscape architects than any other organization, public or private. It is worthwhile, then, to recognize an important deficiency in their relationship with their clients.

In private practice, the client-architect relationship is intense; there is considerable interplay as the designer seeks to incorporate the client's desires into a design reality. But in the public sector this basic relationship is essentially absent, especially with regard to esthetic preferences. If "beauty is in the eye of the beholder", public esthetic perceptions ought to be an important consideration for the landscape designer. Accurate determination of these preferences is therefore essential if the client-architect relationship is to be a reality in the context of public land management.
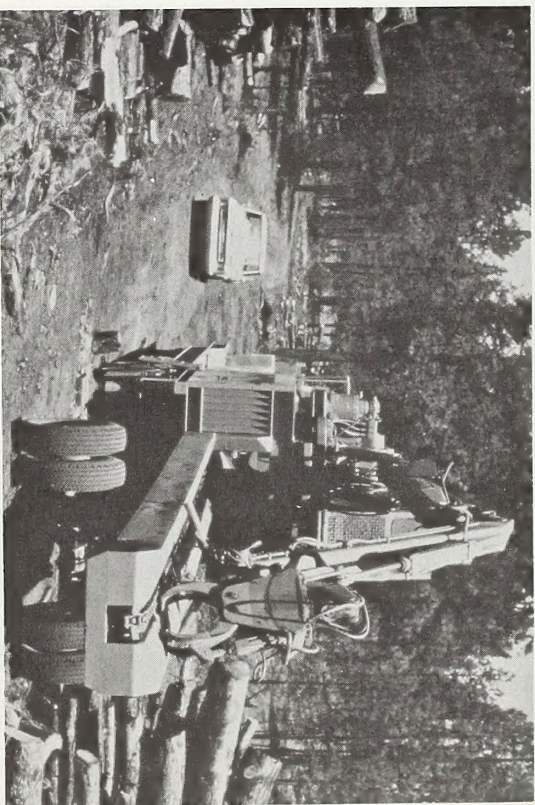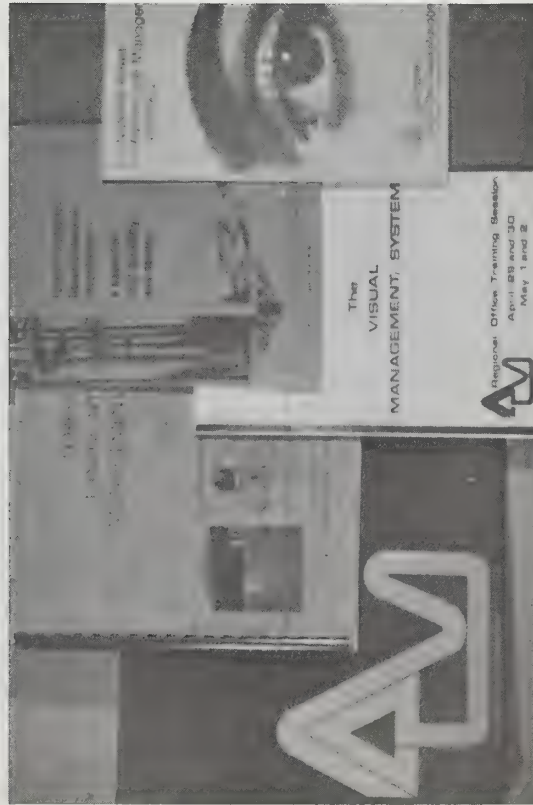
Approaches to the Problem

Public demand for all forest resources has risen substantially in recent years, primarily as a result of dramatic economic growth and, to a lesser degree, because of increased population. Greater affluence and increased leisure time have created a particularly high demand for the less tangible resources: wildlife, recreation opportunity, and scenic beauty. Although the importance of scenic resources has been recognized for some time, attempts at systematic assessment of landscape scenic beauty have been concentrated in the past 10 years. In this short period a variety of schemes and techniques for describing, inventorying, and evaluating scenic quality have been proposed. Methods to date range from highly subjective, general descriptive procedures to more complex quantitative procedures (Boster, in press). Substantial progress has been made, but no method for determining the scenic beauty of landscapes has yet been generally accepted.

Scenic quality assessment methods may be divided into three general categories: (1) descriptive inventories, (2) surveys and questionnaires, and (3) evaluations of perceptual preference. Each of these approaches will be described and their respective advantages and disadvantages discussed. For a more detailed review see Arthur and Boster (1976), or the reviews by Fabos (1971) and Redding (1973).

Descriptive Inventories

Inventories have been extensively used as a means of representing and evaluating landscape quality, but specific methods vary greatly. Some are highly subjective and provide only amorphous lists of vaguely defined landscape features (such as "warmth," "variety," "harmony") that are, at best, intuitively related to scenic beauty. More sophisticated inventory methods have been developed recently; these are less subjective and provide structured lists of relatively well-defined and sometimes numerically scaled landscape features that are formally related to scenic beauty.

Basically, the inventory approach requires that a set of landscape features or components, thought to be relevant to scenic beauty, be selected and, to some extent, defined. Assessment of a given land-

scape involves taking an inventory of the selected features. The presence or absence of each feature is noted, their number counted, and in some instances, a numerical value assigned to each feature. The final step is to relate the feature-based inventory to scenic quality. In some instances this has been accomplished by summing or otherwise combining the features or their numerical values (if determined) to produce a scenic beauty index. An example is the method described by Leopold (1969). Usually, the relationship of the descriptive inventory to scenic quality is presumed or left to individual discretion.

The most widely used and probably the most successful descriptive inventory methods have been developed by landscape architects. The pioneering work of Burton Litton (1968) may be credited for much of the impetus in the development of this approach. Litton-based methods generally emphasize the identification and evaluation of land forms. Representative elements of the landscape typically used in these systems are **line, texture, contrast, and color.** The most recent and comprehensive statement of this approach, and a review of some applications to specific landscape assessment problems, can be found in two recent Forest Service publications (USDA 1973, 1974).

Descriptive inventories offer several advantages as a method of assessing scenic quality. If an appropriate list of relevant features can be established, a wide variety of landscape types can be described in comparable terms. An obvious difficulty is that the features used must be sufficiently flexible to allow application to a number of different landscapes. At the same time, however, they must be specific enough to allow reliable discrimination among a variety of landscapes.

An important advantage of the inventory approach is that it permits landscape evaluations on a large scale. Descriptive inventories have been applied to characterize entire geographic regions encompassing everything from wilderness to cities. The same basic technique has also been applied to describe in detail small, individual landscape scenes. Litton (1974) has noted that differences of scale relate not only to detail, but to whether the intention is to serve **planning** or **design.** One cost of this expansive range of applicability is that distinctions among landscapes tend to be gross (such as natural versus urban landscapes). While large differences in scenic quality might be
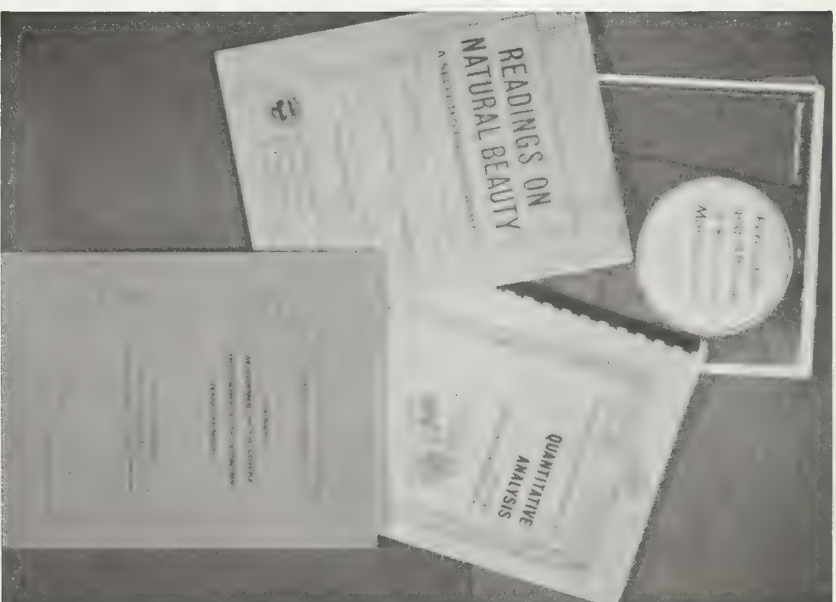
revealed through such a system, few distinctions are likely in the intermediate ranges where most realistic management alternatives lie.

Descriptive inventories will no doubt continue to be an important tool in evaluating the scenic beauty of landscapes. These methods have proven their usefulness, and they are certainly to be preferred over entirely intuitive approaches. However, the effectiveness of inventory techniques depends to a great extent on the expertise and judgment of the user and on the relevance of the descriptive features selected. Perhaps the greatest weakness in this approach has been the failure to relate the features in the inventory to validated measures of scenic beauty.

Surveys and Questionnaires

Questionnaires or opinion surveys have been widely used as a means of determining the desirability of various management alternatives (Potter and Wagar 1971). Survey results can provide an evaluation or assessment of landscape quality by indicating the preferences of those sampled. An important, though often not explicit, underlying assumption is that the respondents' expressed preferences are related to landscape beauty. In the simplest case, the relationship is assumed to be direct, the greater the preference the higher the beauty.

In general, the survey method requires that a question or set of questions relevant to landscape quality be drafted and presented in written or oral form to selected respondents. Very broad, general questions may be posed that require an open or free-form response — for example, "What do you feel are the essential components of a beautiful landscape?" At the other extreme, very specific questions may be posed and the response format may be highly structured, allowing only a multiple choice or graded response. For example, the statement, "Small clearings in the forest enhance scenic beauty of the landscape", may be paired with the graded response alternatives, "strongly agree, moderately agree, moderately disagree, strongly disagree." Responses are usually collated and analyzed to yield summary indications of opinions and preferences of respondent groups. Subsequently, group preferences may be related to landscape beauty.

Written descriptions have been especially popular as a means of representing the effects of management alternatives on intangible resources. In many instances this would appear to be the only viable technique for representing the relevant aspects of a management decision. However, it is important to be aware of the hazards inherent in this procedure.

The specific wording chosen to represent a landscape or landscape feature must be selected from a very large number of seemingly parallel alternatives. The specific wording is critical, however, because responses are often determined as much by the way a landscape is represented as by its actual characteristics. The statement above about the scenic effect of "small clearings in the forest" provides a striking example. As posed, the statement generally produces strong positive responses. A parallel statement, "Small clearcut patches in the forest enhance scenic beauty of the landscape" could produce quite the opposite result.

There is no simple way to determine the best phrasing for questions to be used in survey instruments. A similar problem arises in trying to determine the "true meaning" of respondents' replies, especially when open response formats are used. For example, surveys have generally indicated strong public preferences for "natural" landscapes. When the same public is asked to indicate their preferences for actual (but unlabeled) landscapes, however, they frequently prefer intensively managed forest on the basis of greater "naturalness."

Surveys and questionnaires are often employed because they are an efficient and economical way to sample the reactions of a large segment of the public. Printed forms can be reproduced cheaply and distributed quickly. However, a classic problem can occur when some percentage of the questionnaires are not returned. The surveyor may take great pains to send his forms to a "representative sample," only to have his true sample determined by the willingness of individuals to fill out and return the form. This problem is compounded by the fact that the length and format of the questionnaire, the general topic

Similar problems arise in trying to determine the "true meaning" of respondents' replies, especially when open response formats are used. For example, surveys have generally indicated strong public preferences for "natural" landscapes. When the same public is asked to indicate their preferences for actual (but unlabeled) landscapes, however, they frequently prefer intensively managed forest on the basis of greater "naturalness" landscapes, however, they frequently prefer intensively managed areas (Daniel et al. 1973, Zube 1974). Moreover, observers frequently justify their preference for a carefully designed park or intensively managed forest on the basis of greater "naturalness."

being surveyed, and even the return address may influence who returns the form and, perhaps, how the form is filled out (White 1975).

While surveys do have many shortcomings, they remain a useful tool for determining public preferences. The hazards of this approach must be noted, however, and appropriate care taken in the construction, distribution, and interpretation of survey instruments.

Perceptual Preference Assessment

Perceptual preference approaches are similar in many respects to the questionnaire or survey procedures. Both evaluate landscape quality through the judgments of human observers. These methods explicitly recognize that observer judgments are especially relevant for evaluating scenic quality.

Perceptual judgment procedures generally represent the landscape more directly than do verbal surveys. Photographs and other graphic representations are used extensively. Less often, the actual landscape is visited for evaluation. Usually, judgments are collected directly, eliminating the "no return" problem encountered with mailed questionnaires.

While photographs or color slides are often used to represent landscapes, methods for selecting and presenting them vary among specific applications. Aside from determining vantage points, a decision must be made as to which photo or photos should represent a particular landscape. If scenes are selected on the basis of professional photographic criteria (composition, framing, color, etc.) or if the evaluator selects representative scenes, substantial bias may be introduced which can affect judgments. Another approach is to use systematic, random sampling. Depending upon the size and variability of the landscape, this method may require that a larger number of photos be taken to insure adequate representation. In some situations, random sampling may be inappropriate (where prominent features are present or where vistas tend to attract attention). Regardless of the method for selecting photos, the primary concern must be an accurate representation of the actual landscape. Studies have shown (Boster and

Vantage point, what Litton (1968) calls "observer position", is an important consideration in any attempt to measure scenic beauty. Some areas are seen primarily as panoramic vistas from miles away; others are primarily seen from a jet at 40,000 feet; many others are most typically viewed from a car at various speeds; still others are viewed on-site. Of course, most areas are viewed in several ways. This variation in vantage point can cause problems. For example, clearcuts, when viewed from afar, may have markedly different scenic beauty than when viewed on-site. Vantage point, then, raises the question of how best to represent a scene. Specifically, how and from where should the pictures be taken?

Vantage Point

9

Daniel 1972, and in this paper; Zube 1974) that color slides or photographs can represent actual landscapes quite well.

A number of procedures have been used for showing the photographic representations to observers and obtaining their judgments. While some methods have required judgments regarding specific characteristics or components of landscape quality (color, contrast, "naturalness"), most have directly asked for judgments of esthetic preference or scenic quality. Observers may be asked, in a forced-choice procedure, to choose the "best" of two or more landscape scenes. Alternatively, an ordinal ranking of several scenes may be required, or the observer may be asked to assign a specific value (rating) to each of several scenes separately. Generally, a relatively large sample of observers is used and their individual scenic quality judgments are combined to produce a single index of group judgment. The group index (e.g., percent choice, average rank, or mean rating) may be used directly as an estimate of the scenic beauty of the evaluated landscape.

Each judgment procedure has advantages and disadvantages. Forced-choice and ranking procedures require the observer to distinguish among the landscapes being evaluated, but they generally provide only an ordinal ranking with little indication of relative preference intensities. Further, only a small number of scenes can be effectively managed with either of these procedures. For example, to evaluate ten landscape scenes in a pair-wise, forced-choice procedure, a minimum of 45 pairs of scenes must be presented; ranking procedures require that all of the scenes be available simultaneously, and few observers could be expected to judge and rank more than 10 scenes at one time. By using an individual rating procedure, a much larger number of landscapes can be efficiently evaluated and an indication of relative differences between landscapes (rather than simple rankings) can be obtained. Proper interpretation of rating responses, however, often requires mathematical transformations and/or statistical analysis to adjust for observer's idiosyncratic use of the response scale.

Perceptual evaluation approaches are among the most recent developments in the effort to assess scenic beauty. As a consequence, they

Standardization of Ratings

Many investigators have recognized the problem created when observers adopt different strategies for assigning rating values. The most common procedure for dealing with this problem is to transform individual observer's ratings to standard (z) scores by the general formula:

$$z_{ij} = \frac{R_{ij} - \bar{R}_j}{s_j}$$

where: z_{ij} = standardized (z) score for the ith rating response of observer j.

$\bar{R}_j$ = mean of all ratings by observer j

R_{ij} = i^{th} rating of observer j

s_j = standard deviation of all ratings by observer j.

The z transformation, then, produces a scale with an origin equivalent to the mean of the observer's ratings (i.e., the mean rating, $\bar{R}_j$, will be transformed to zero on the z-score scale). Further, all values on the z scale will be expressed in units equal to the standard deviation of the observer's ratings; that is, if the observer's standard deviation were 3.0, a difference of 3 on the rating scale would be a difference of 1.0 on the z-transformed scale. Thus, arbitrary differences between observers in how they use the rating scale, both in terms of tendencies to use only the high or low end of the scale and differences in the extent or range of the scale used, would be eliminated by the z transformation. Also, results of different scales (such as 5-point or 7-point) could be directly compared.

There are important problems with this conventional standardization approach, however, when it is applied across different observers' ratings of areas that may actually differ in scenic beauty. First, using each observer's mean rating averaged across all evaluated areas as the origin for his transformed scale may obscure real differences between individual observer's judgments. The same mean rating may be obtained for two observers, even though one assigns medium range ratings (4-6) to all landscapes and the other assigns very low range ratings (1-3) to some landscapes and very high ratings (8-10) to others. For the latter observer, transformed values will all tend to be rather extreme (high positive vs. high negative z values) and apparent dif-

ferences in the values at either end of the scale may be excessively reduced.

Conventional z transformations tend to compound the origin problem by dividing by the overall standard deviation of rating responses (thus adjusting the size of the unit). The standard deviation is usually computed across all ratings, irrespective of the different landscape areas being judged. This procedure fails to distinguish between two different sources of variance in the ratings: variability in ratings assigned to repeated instances from the same landscape area (error, landscape heterogeneity) and variability between areas (discrimination of differences in landscape beauty). Conventional standardization fails to distinguish, for example, between observers who spread their ratings equally between and within landscapes and observers who produce "tight" distributions within categories, but distinguish sharply between categories. The origin problem discussed above would tend to be further exaggerated, then, because the observer assigning all medium ratings would have a lower overall standard deviation than the observer who distinguished sharply between the landscapes. Thus, apparent differences between the second observer's z values (especially near the extremes of the scenic beauty scale) would be even further reduced.

Inappropriate standardization procedures can indicate differences in observer reactions (perceptions) where there are only idiosyncracies in rating-scale use. An equally important hazard is that true differences in observer reactions may be obscured by using transformations that fail to distinguish variations in ratings within one area from systematic variations in ratings between areas.

have not been studied or applied as extensively as the questionnaire/ survey and descriptive inventory procedures. The explicit incorporation of public responses into the evaluation procedure is an important feature, particularly where assessment of public lands is concerned. Because they represent the landscape more directly, they offer distinct advantages over surveys.

LANDSCAPE EVALUATION: THE SBE MODEL

The first question that must be considered in developing any measurement system is: "What is to be measured?" Answering this question proves particularly difficult when the objective is to appraise scenic beauty. The problem of defining "beauty" has occupied the minds and pens of philosophers for centuries, and we are under no delusion that we have laid the problem to rest. However, it is essential to have an explicit conceptualization of scenic beauty to guide the development of a measurement system.

Conceptual Approach

Our conceptualization of scenic beauty is based on the premise that beauty is an "interactive" concept. Scenic beauty is neither entirely "in the eye of the beholder" nor is it solely a property of the landscape. Rather, scenic beauty is inferred from a judgment made by a human observer in response to his perception of a landscape.

This conceptualization is represented by the dual component model of scenic beauty illustrated by the adjacent figure. Evaluative judgments are represented as a combined product of the observer's perception of scenic beauty in the landscape (Landscapes 1 and 2 in the figure) and his judgmental (esthetic) standards or criteria (the arrow in the figure).

Because the perceived beauty of Landscape 1 falls short of the observer's minimum criterion for landscape beauty, a negative judgment ("I don't like it" or "it is ugly") will result. The perceived beauty of Landscape 2, however, exceeds the observer's criterion and a positive judgment results ("I like it" or "It is beautiful"). Should the observer's standards be raised for some reason (to point A, for example) his judgment would be negative for both landscapes, **even though their perceived beauty has not changed.** Thus, scenic beauty judgments depend jointly on the perceived properties of the landscape and the judgmental criteria of the observer.
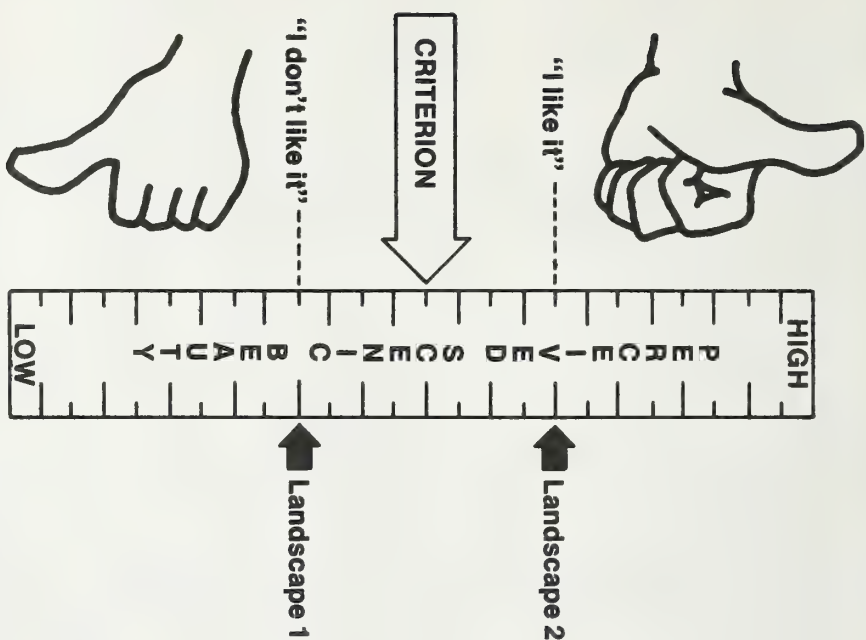
Perceptual Component

The perceived scenic beauty of a landscape is represented in the model by a value on a perceptual dimension. Each value is assumed to result from the combined effects of a number of visible landscape properties. Some progress has been made in identifying scenically relevant physical features (Shafer, Hamilton and Schmidt, 1969, USDA 1973, 1974), but neither the features nor the "rules" by which perceptual effects combine can yet be accurately specified.

If one area is consistently judged more beautiful than another, the areas are assumed to have different values on the perceived scenic beauty scale. As illustrated above, however, changes in criteria may produce different judgments even though the perceived scenic beauty is the same. To determine the effects of perceived differences in landscape beauty, the judgment criteria of the observer must also be taken into account. If observer criteria may be assumed constant and judgments of landscapes differ consistently, then a difference in perceived scenic beauty is indicated.

Judgment Criterion Component

The criterion component of the SBE Model is also represented by a value (or set of values) on the perceived scenic beauty dimension. These criterion values, brought to the judgment situation by the observer, are important components of any evaluative judgment. If the observer is required to make a binary judgment (either "I like it" or "I don't") he need only establish a single criterion value (as illustrated by the arrow in the figure). Any scene producing a perceived scenic beauty value above the criterion will result in an "I like it" evaluative judgment. Perceived values below the criterion will result in "I don't like it" judgments. In more complex judgment situations the observer may have to establish several criterion values, each determining one of several possible evaluative judgments ("I'll pay $10 but no more," or "I rate that a 6 on a 10-point scale").

For scenic beauty judgments of forest landscapes, observers may have different criteria depending upon the nature of their past experiences with forests. An observer whose experience has been restricted to mesquite trees and cactus, for example, may be very complimentary of any forested landscape. The same landscape, however, may be judged as rather unimpressive by an observer who grew up in the California redwoods.

The immediate context in which an observer judges a landscape may also influence the criteria that are applied. When a friend shows you a site that he recently purchased for a summer home, the tendency is to be complimentary. Were the same site presented by a salesman trying to sell you the lot, you might be more cautious and reserved in

15
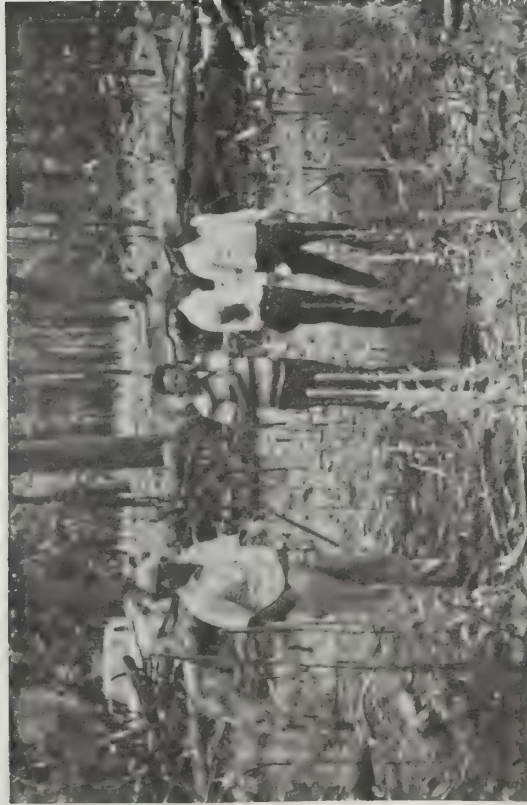
offering praise. In general, whenever observers consistently rate the same landscape differently (as when one observer assigns a "3" and the other an "8"), it is likely that different criteria are being applied; their **perceptions** may or may not be identical. The perceived scenic beauty of the landscape and the observer's criteria jointly determine any judgment of landscape beauty, but neither the perceptual nor the criterion component can be assessed directly. The separate effects of each component can be determined, however, by a systematic analysis of evaluative judgments of different landscapes.

Analytical Approach

An observer may express his judgments of landscape beauty by a numerical rating. Each landscape scene can be assigned a rating from a scale — for example, 1 (extremely low scenic beauty) to 10 (extremely high scenic beauty). The use of a rating scale requires the observer to establish several separate, ordered criterion values.

The criterion for offering a rating of "1" would be lower than that for a "2", and the highest value (most stringent criterion) would be reserved for a "10" judgment. Thus, when an observer's rating of a particular landscape scene is "7", he indicates that the perceived beauty for the scene has reached or exceeded his criterion for a "7" judgment, but fallen short of the value required for a rating of "8".

Different observers may establish different criteria for assigning ratings. Because the observer's esthetic standards cannot be directly determined, ratings by themselves can be difficult to interpret. A rating of 3 may indicate a high degree of perceived scenic beauty if the observer is applying very high (stringent) esthetic criteria. On the other hand, a rating of 8 may indicate rather low landscape beauty when assigned by an observer having low (lax) esthetic standards.

Differences in ratings may indicate true differences in perceived scenic beauty. If an observer rates one landscape scene a 3 and another an 8, the two landscapes must have different scenic values (unless his criteria suddenly shifted). A problem arises here, however, when we attempt to determine how much of a difference in scenic beauty is indicated by these ratings. A different observer might, for example,

More on the Hazards of Ratings

Attaining a quantitative measure of the scenic beauty of a forest landscape based on numerical ratings assigned by different observers (or groups of observers) can pose a number of problems. The most direct approach would seem to be to gage the scenic beauty of the landscape in terms of the actual ratings that have been assigned to it — using the average (mean) rating, for example, as a beauty index. Unfortunately, this simple and straightforward procedure has some important inadequacies. First, there is the obvious problem that such an index would have to be interpreted differently depending upon whether a 10-point, 5-point, or some other scale was used. Clearly, an average rating of 5 would be very "good", if only a 5-point scale was used. Further, the range of the scale has important implications for interpreting differences in indices — a difference of 2 points would be more significant on a 5-point scale than on a 10-point scale.

Aside from the obvious disadvantages of having an index that must be interpreted differently for different applications, there are other, more subtle difficulties to be noted. Tabled and graphed below are some hypothetical data to illustrate some of these problems. Mean ratings (10-point rating scale) are presented for four landscapes as they might have been assigned by three different observers:

	Observer		
Landscape	A	B	C
I	3	5	2
II	1	3	1
III	8	10	5
IV	5	7	3

The observers' reactions to the landscapes appear to differ considerably. If the mean ratings were taken as a measure of scenic beauty, we would conclude that Observer A, and that Observer B finds each landscape more beautiful than do either of the other observers. However, there is no guarantee that all three observers are using the 10-point rating scale in the same way — their judgment criteria may be different. Inspection of the ratings assigned by Observers A and B reveals that their judgments differ by a constant amount; their relative ratings of the landscapes are actually the same. By making each observer's

rate the first scene a 5 and the second a 6. Does the 5-point spread in judgments given by the first observer indicate a greater difference in scenic beauty than the 1-point difference in the second observer's ratings?

The ambiguity introduced by differences in observers' criterion values can be eliminated. Briefly, the Scenic Beauty Estimation (SBE) procedure provides measures of landscape beauty **independent** of observer judgmental criteria. We term these values **Scenic Beauty Estimates**, or SBEs for short. SBEs are relative scale values calculated from each observer's ratings of a number of different landscape scenes. The contention that beauty is an interactive concept is directly represented in the SBE Model. Measurements of scenic beauty are not derived entirely from characteristics of the landscape, nor are the stated preferences of observers (ratings) assumed to be a direct index of perceived beauty. Scenic beauty values are derived from judgments made by a number of observers for a variety of landscapes. Observers' ratings are adjusted to take into account the effect of differing judgment criteria. The resulting **Scenic Beauty Estimates** (SBEs) provide a quantitative index of the perceived scenic beauty of the landscapes.

ratings relative *to his ratings of one of the landscapes (by subtracting his mean rating of landscape I from each other landscape's mean rating) the apparent differences among the observers' evaluations are reduced greatly, as indicated in the table and graph below:*

	Observer		
	A	B	C
Landscape			
I	(3-3=) 0	(5-5=) 0	(2-2=) 0
II	(1-3=) -2	(3-5=) -2	(1-2=) -1
III	(8-3=) 5	(10-5=) 5	(5-2=) 3
IV	(5-3=) 2	(7-5=) 2	(3-2=) 1

In fact, Observers A and B now produce identical sets of indices for the four landscapes. The indication is that these two observers differed only in terms of their relative standards or judgment criteria, and not necessarily in their perception of the scenic beauty of the three landscapes. If this were the only difference between observers (and identical rating scales were being used), this simple difference or original adjustment procedure would be adequate to provide unbiased indices of the perceived scenic beauty of the landscapes. However, as is illustrated by the data for Observer C, criterion effects may be more complex.
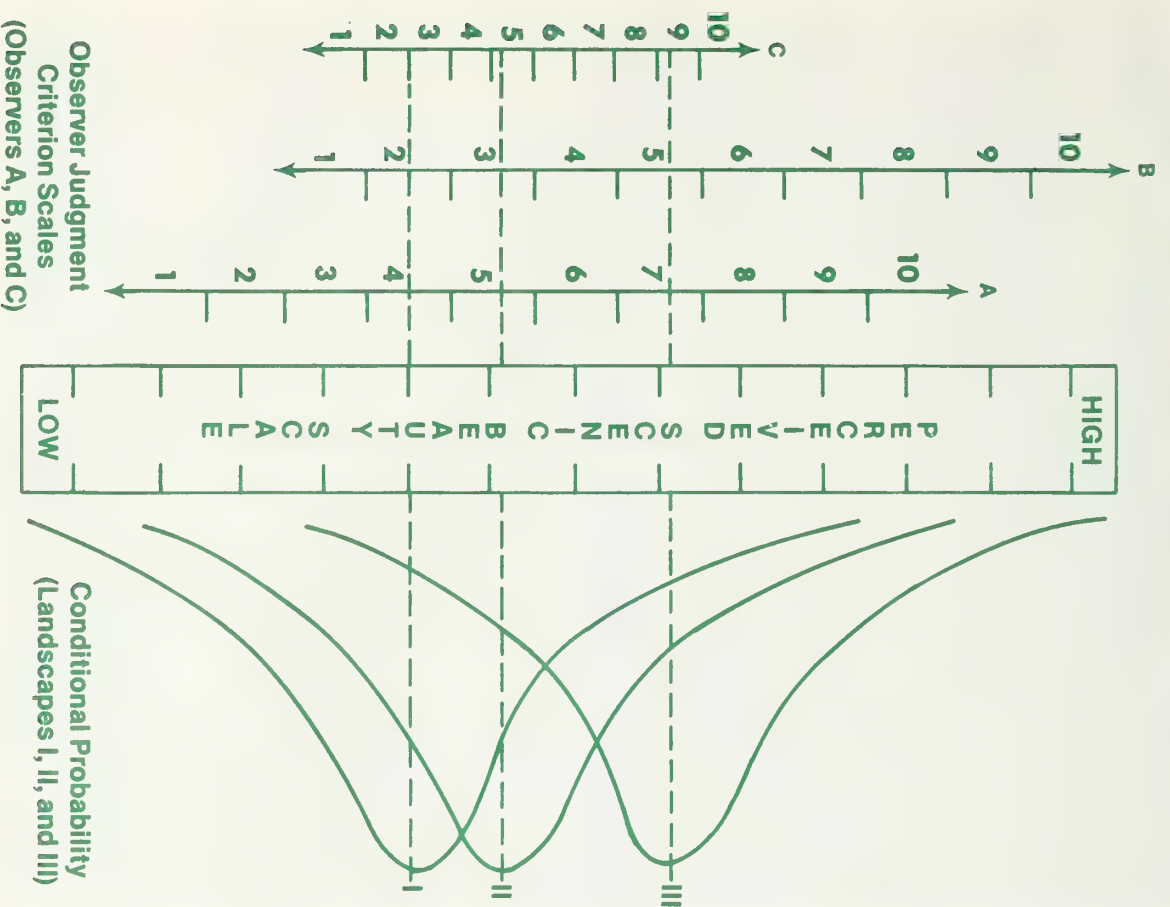
Observer C may actually perceive the landscapes as being less different in scenic beauty — as suggested by the graph — or, he may have consistently cut off one or more points on either or both ends of the 10-point scale, rather than using the entire range of the scale. To distinguish between these possibilities, differences in the **variability** *of ratings (from one observer to the next) must be taken into account. This problem is more complex than the simple relative difference situation between Observers A and B.*

Some aspects of this "unit size" problem were discussed above under "Standardization of Ratings". Under "The SBE Model", below, the way in which the SBE computational procedures deal with both the simple, relative differences and the more complex unit size differences are presented in detail. For now, it need only be noted that these two types of problems do occur. The SBE Method avoids these problems by providing appropriately standardized measures (SBEs) that are unaffected by observer criterion differences, and for the most part, are not affected by the size of the rating scale used.

SBE Model: Hypothetical Example

Observer Judgment
Criterion Scales
(Observers A, B, and C)

Conditional Probability
(Landscapes I, II, and III)

The SBE Model

The ratings assigned to a landscape are determined by the relation-ship between the perceived scenic beauty of the landscape and the judgmental criteria being applied by the observer. This conceptual approach is most explicitly suggested by the Theory of Signal Detect-ability as formulated by Green and Swets (1966, also Swets 1973). Similar conceptualizations may be derived from classical phycho-physical scaling procedures developed by Thurstone (1927, 1948) and discussed by Torgerson (1958) and Bock and Jones (1968). An excellent and very readable discussion of these and other classical procedures may be found in Hays (1967). Some relationships between signal detection methods and Thurstone's scaling procedures are discussed by Lee (1969). The SBE Model and analytical procedures and include aspects of both the signal detectability model and Thurstone-type scaling models. The basic components of the SBE Model are briefly described below.

The perception of a landscape is represented by a distribution of "perceived scenic beauty" values. The distribution reflects the variabil-ity of perceptual effects produced by different scenes sampled from the same landscape and by momentary fluctuations in the observer's perceptual processes. Thus, the "perceived scenic beauty" of a land-scape is not considered to be a single value, but may be represented by the average of the many perceived values that may result from a number of scenes from that landscape.

It is assumed that if the scenes representing a landscape are sampled in an unbiased fashion (i.e., randomly), the perceived value distribution for the landscape will be distributed normally. Hypothetical distribu-tions of "perceived scenic beauty" values for three landscapes are presented for illustration in the adjacent figure.

Note that there is considerable overlap among the distributions of the first two landscapes. The implication is that, while landscape I (LS-I) is generally lower in scenic beauty than LS-II, occasionally scenes from LS-I will be perceived as more beautiful than some scenes from LS-II.

Three hypothetical observer-criterion scales are presented to the left of the landscape distributions. Each observer's criterion values divide the perceived scenic beauty scale into 10 segments. According to the

Frequencies (f), cumulative frequencies (cf), cumulative probabilities (cp), Z scores (z), Mean Zs ($\bar{z}$), for three landscapes

Observer A

Rating	LI f	LI cf	LI cp	LI z	LII f	LII cf	LII cp	LII z	LIII f	LIII cf	LIII cp	LIII z
1	2	30	1.00	—	1	30	1.00	—	0	30	1.00	—
2	3	28	.93	1.47	2	29	.97	1.84	1	30	1.00	2.13
3	5	25	.83	.95	3	27	.90	1.28	1	29	.97	1.84
4	8	20	.67	.44	5	24	.80	.84	2	28	.93	1.47
5	5	12	.40	-.24	8	19	.63	.33	3	26	.87	1.12
6	3	7	.23	-.74	5	11	.37	-.33	5	23	.77	.74
7	2	4	.13	-1.12	3	6	.20	-.84	8	18	.60	.25
8	1	2	.07	-1.47	2	3	.10	-1.24	5	10	.33	-.44
9	1	1	.03	-1.84	1	1	.03	-1.84	3	5	.17	-.95
10	0	0	0.00	-2.13	0	0	0.00	-2.13	2	2	.07	-1.47

Landscape I: $\Sigma z = -4.68$, $\bar{z} = -.52$, SBE = [-.52 - (-.52)] x 100 = 0
Landscape II: $\Sigma z = -2.19$, $\bar{z} = -.28$, SBE = [-.28 - (-.52)] x 100 = 24
Landscape III: $\Sigma z = 4.69$, $\bar{z} = .52$, SBE = [.52 - (-.52)] x 100 = 104

Observer B

Rating	LI f	LI cf	LI cp	LI z	LII f	LII cf	LII cp	LII z	LIII f	LIII cf	LIII cp	LIII z
1	10	30	1.00	—	6	30	1.00	—	1	30	1.00	—
2	8	20	.67	.44	5	24	.80	.84	2	29	.97	1.84
3	5	12	.40	-.24	8	19	.63	.33	3	27	.90	1.28
4	3	7	.23	-.74	5	11	.37	-.33	5	24	.80	.84
5	2	4	.13	-1.12	3	6	.20	-.84	8	19	.63	.33
6	1	2	.07	-1.47	2	3	.10	-1.28	5	11	.37	-.33
7	1	1	.03	-1.84	1	1	.03	-1.84	3	6	.20	-.84
8	0	0	0.00	-2.13	0	0	0.00	-2.13	2	3	.10	-1.28
9	0	0	0.00	-2.13	0	0	0.00	-2.13	1	1	.03	-1.84
10	0	0	0.00	-2.13	0	0	0.00	-2.13	0	0	0.00	-2.13

Landscape I: $\Sigma z = -11.36$, $\bar{z} = -1.26$, SBE = [-1.26 - (-1.26)] x 100 = 0
Landscape II: $\Sigma z = -9.51$, $\bar{z} = -1.06$, SBE = [-1.06 - (-1.26)] x 100 = 20
Landscape III: $\Sigma z = -2.13$, $\bar{z} = -.24$, SBE = [-.24 - (-1.26)] x 100 = 102

Observer C

Rating	LI f	LI cf	LI cp	LI z	LII f	LII cf	LII cp	LII z	LIII f	LIII cf	LIII cp	LIII z
1	9	30	1.00	—	6	30	1.00	—	2	30	1.00	—
2	4	21	.70	.52	2	24	.80	.84	1	28	.93	1.47
3	3	17	.57	.16	3	22	.73	.61	1	27	.90	1.28
4	3	14	.47	-.08	4	19	.63	.33	2	26	.87	1.12
5	3	11	.37	-.33	4	15	.50	.00	2	24	.80	.84
6	2	8	.27	-.61	3	11	.37	-.33	3	22	.73	.61
7	2	6	.20	-.84	2	8	.27	-.61	3	19	.63	.33
8	1	4	.13	-1.12	2	6	.20	-.84	3	16	.53	.08
9	1	3	.10	-1.28	1	4	.13	-1.12	4	13	.43	-.16
10	2	2	.07	-1.47	3	3	.10	-1.28	9	9	.30	-.52

Landscape I: $\Sigma z = -5.00$, $\bar{z} = -.56$, SBE = [-.56 - (-.56)] x 100 = 0
Landscape II: $\Sigma z = -2.40$, $\bar{z} = -.27$, SBE = [-.27 - (-.56)] x 100 = 29
Landscape III: $\Sigma z = 5.05$, $\bar{z} = .56$, SBE = [.56 - (-.56)] x 100 = 112

Note: Because of the cumulation process, there will necessarily be a cp = 1.00 at the lowest rating level (1), so that no information would be added by a z-score entry identical for every observer. The $\bar{z}$ is therefore calculated on the basis of one less than the number of rating responses (10 − 1 = 9 in this case). In other cases where cp = 1.00 or where cp = 0.0 (z = ± ∞), we have adopted the convention of cp = 1 − 1/(2N) or cp = 1/(2N), respectively (see Bock and Jones 1968). In the above example, N = 30, hence, cp = 1 − 1/60, yielding z scores of 2.13 and − 2.13, respectively.

SBE model, the ratings assigned by an observer to any given landscape depend upon the judgment criteria established. Thus, scenes sampled from LS-I will most often be given ratings of 4 by Observer A. Observers B and C on the other hand, will more often assign lower ratings (1, 2, or 3). These differences in ratings occur in spite of the fact that, for this hypothetical example, the perceived scenic beauty values are identical for all three observers.

Ratings often differ, of course, because of true differences in perceived scenic beauty. To illustrate true perceptual differences, judgments of more than one landscape must be considered. Consistent differences in the ratings assigned to several landscapes by the same observer will reflect separation among those landscapes on the perceived scenic beauty scale. If, for example, the distributions of values for two landscapes overlapped completely, there would be no difference in the distribution of ratings assigned for those two landscapes, regardless of the nature of the observer's judgment criteria. To the extent that an observer's ratings for two landscapes are consistently different, the landscapes must have different perceived scenic beauty distributions.

The table presents hypothetical rating distributions for Landscapes I, II, and III for Observers A, B, and C. The tabled values are related to the SBE "example" figure in that the rating distributions approximately reflect the area under the respective perceived-beauty distributions covered by each observer's 10 rating-criterion categories. The frequency (f), cumulative frequency (cf), and cumulative probability (cp) values are shown for each landscape. The z values are the standard normal deviates associated with each of the cumulative probability values. Means and standard deviations for the z values are also shown. These tabled values may be related to the Hit Rate and False Alarm Rate tables of signal detection methods (Green and Swetts 1966, Hake and Rodwan 1966) or to the "M" and "X" matrices of classical psychophysical scaling procedures (Torgerson 1958, Lee 1969).
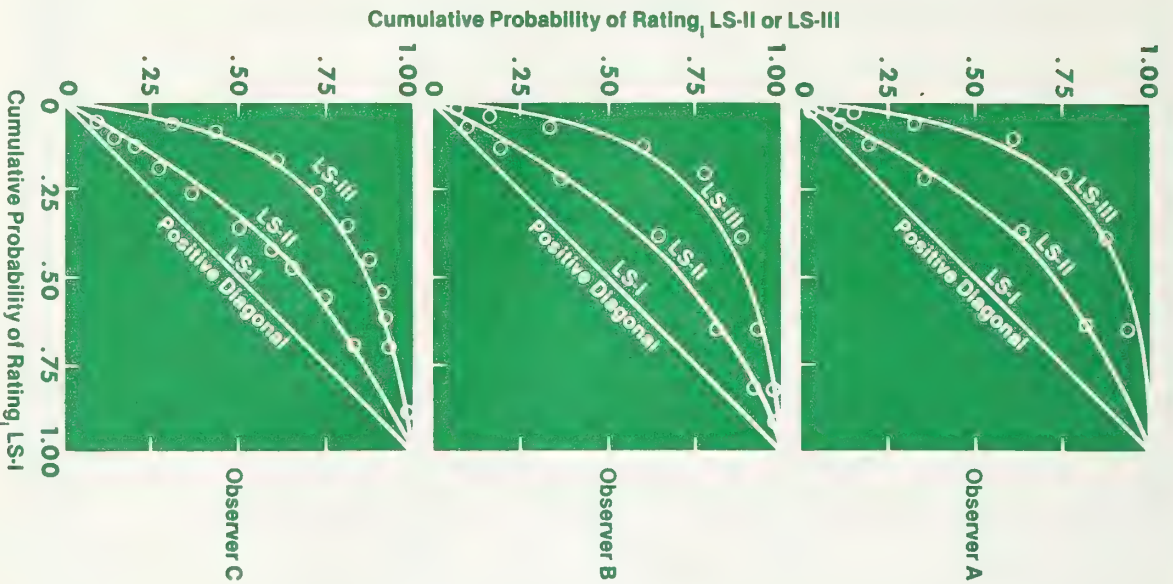
Our SBE analysis assesses differences in perceived scenic beauty by comparing an observer's rating distribution for one landscape area against each of several others. This may be accomplished graphically by plotting a relative operating characteristic (ROC), a bivariate graph of the cumulative probability of the ratings (1-10) for the selected comparison landscape, against the cumulative probabilities of the ratings (1-10), respectively, for each of the other landscapes.

Hypothetical Relative Operating Characteristics (ROCs)

If LS-I is selected as the basis for comparison, and cumulative probabilities for LS-II and LS-III are plotted separately against those for LS-I at each rating category, two relative operating characteristics (ROCs) result for each observer. The frequency tables for an observer are based upon his ratings of the randomly sampled scenes (N = 30 in this example) representing the respective areas (landscape I, II, and III).

Notice that for each observer as the difference between the cumulative probability of a rating for LS-I (absisa) and the cumulative probability of that same rating for the other landscape (ordinate) increases, the distance of the ROC from the positive diagonal increases. If there were no difference between perceived scenic beauty values for two landscapes (for example, if LS-I values were compared to themselves) the ROC would exactly follow the positive diagonal. The ROC for LS-III is generally farther from the diagonal than is that for LS-II reflecting the fact that, in our hypothetical example, perceived scenic beauty values for LS-III are generally higher than those for either LS-I or LS-II. The distance of the ROC from the positive diagonal is a measure of the perceived scenic beauty of a landscape relative to the selected comparison landscape (in this case, LS-I).

The ROC functions for the three hypothetical observers are essentially identical in spite of the rather gross differences in their respective judgment criteria. It is in this sense that the ROC is an unbiased (by judgment criteria) representation of perceived scenic beauty of each landscape for each observer.

The most commonly used procedures for determining the distance of the ROC from the positive diagonal require that the cumulative probability values be transformed to standard normal deviates (z scores). If the ROC were plotted on z-score (normal-normal) coordinates (rather than cumulative probabilities) the ROC for LS-II would be a straight line that can be described by:

$$z_{LS-II} = \left(\frac{\sigma_{z_{LS-II}}}{\sigma_{z_{LS-I}}}\right) z_{LS-I} + \bar{z}_{LS-II} - \left(\frac{\sigma_{z_{LS-II}}}{\sigma_{z_{LS-I}}}\right) \bar{z}_{LS-I}$$

The distance of this standardized ROC from the positive diagonal can be measured from the ROC graph in several ways, but it may also be computed directly from the data. Several measures have been used in signal detection applications, but we have generally used the average distance of the ROC from the positive diagonal ($\bar{d}_m = \bar{z}_{LS-II} - \bar{z}_{LS-I}$)

for reasons explained in earlier publications (Angus and Daniel 1974, Wheeler, et al. 1971). Two other measures, $\underline{d}'$ (the distance of the ROC from the positive diagonal at the y-intercept), and $\underline{d}_s$ (the distance to the positive diagonal from the intercept of the ROC and the negative diagonal) are also calculated in the SBE computer program:

$$\underline{d}' = \bar{z}_{LS\text{-}II} - \left(\frac{\sigma_{z_{LS\text{-}II}}}{\sigma_{z_{LS\text{-}I}}}\right)\bar{z}_{LS\text{-}I};$$

$$\underline{d}_s = \frac{2\underline{d}'}{1 + \left(\dfrac{\sigma_{z_{LS\text{-}II}}}{\sigma_{z_{LS\text{-}I}}}\right)}$$

Unless the slope of the ROC $\left(\dfrac{\sigma_{z_{LS\text{-}II}}}{\sigma_{z_{LS\text{-}I}}}\right)$ deviates markedly from the expected value of 1.0, all three values are essentially equal.

The $\underline{d}_m$ measure multiplied by 100 (to remove the decimal point) is what we call the **Scenic Beauty Estimate**, SBE (multiplication by a constant has no effect on the interval relationship among scores). As illustrated in the table, a separate SBE can be obtained for each observer, based upon his judgments of a number of samples (such as slides) from each of several landscape areas. The individual observer values may then be averaged to obtain an SBE for each area represented.
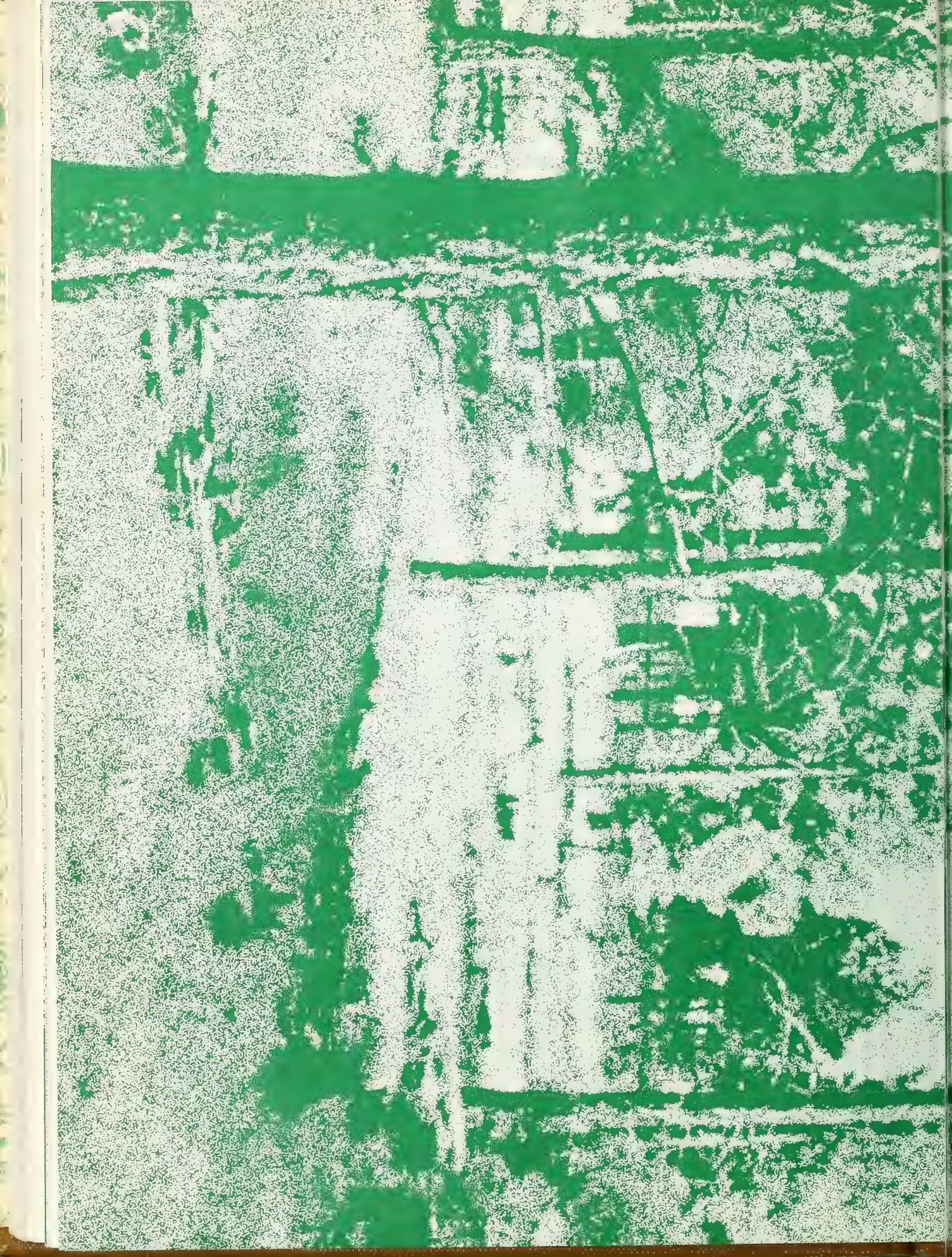
LANDSCAPE EVALUATION: APPLYING THE SBE METHOD

We have applied the SBE Model to evaluate the scenic beauty of forested landscapes. In this section, the basic methods and analytic procedures that we have employed are described and related to the conceptual model. While the model is not restricted to the particular methods we have used, they do illustrate in specific terms how the SBE Model may be implemented.

Application involves three main steps:

1. Representing landscapes by color slides
2. Presenting slides to observers
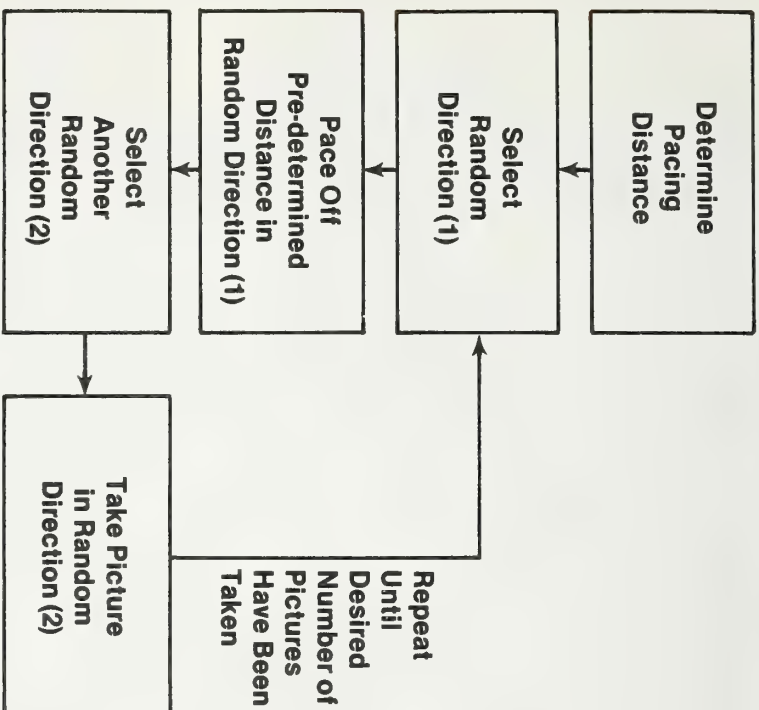3. Evaluating observer judgments

Color prints or even sketches (if properly validated) could also be used. Steps 1 and 2 may actually be eliminated by transporting the observers to the sites; a more direct, but much more costly procedure.

Representation by Slides

To avoid introducing our own biases as to what is "representative", we use an impartial sampling procedure. Areas to be evaluated are first stratified into approximately equal sub-areas. Armed with a compass, a printout of random degrees (1-360), and a camera, the photographer paces across each sub-area following randomly determined directions (bearings), stopping at locations dictated by a pre-determined pacing distance. At each photo-sampling point, a picture is taken at eye level along a randomly determined heading. This procedure is repeated until the desired number of pictures is taken.

The number of slides needed to adequately represent a sampled area and the specific sampling procedures will depend upon several factors and constraints (discussed below). Although the size of an area is important, the number of slides required will depend primarily upon the area's diversity. Relatively homogeneous landscapes might be adequately represented by one slide. For the experiments reported here, 50 slides were taken in each area (10 in each of 5 sub-areas). For our initial experiments, 25 slides per area (5 from each sub-area) were

Slide Sampling Procedure

```
Determine
Pacing
Distance
   ↓
Select
Random
Direction (1)
   ↓
Pace Off
Pre-determined
Distance in
Random Direction (1)
   ↓
Select
Another
Random
Direction (2)
   ↓
Take Picture
in Random
Direction (2)
   →  Repeat
      Until
      Desired
      Number of
      Pictures
      Have Been
      Taken
```

We have found that, from relatively homogeneous observer populations, between 20 and 30 observers is an adequate sample. For some experiments, fewer than 20 observers have provided sufficient sensitivity for statistical comparisons of SBEs among landscape treatments. For all the applications of the SBE Method reported here, the sampled landscapes have been rather large and homogenous (virtually all ponderosa pine forests with relatively flat topography), and the random walk sampling procedure has worked very well. More diverse land-

randomly drawn from the "population" of 50 slides; in some of our later tests, 15 slides (3 per sub-area) were found to be sufficient.
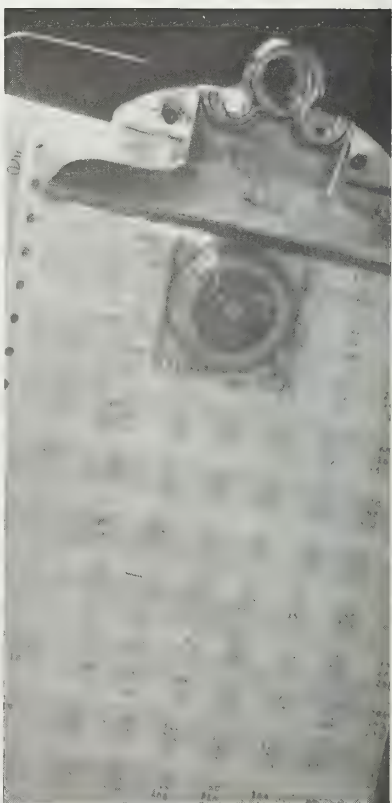
scapes (mixtures of natural and intensively managed elements) and areas with prominent visual features or vistas (that may attract observer attention) might not be best represented by a completely random sampling procedure.

Another sampling procedure that we have tried involves taking pictures at pre-determined angles relative to roads or foot trails. Still another procedure — which we discuss in the final section of this report — is to take photos in randomly determined directions from defined points (such as from survey stakes initially planted for non-esthetic criteria). Other equally valid procedures are possible, but unbiased representation must be insured.

Presenting Slides

Processed slides are labeled, and a stratified random sample of desired size is drawn for each area. These slides are then scrambled into a random order, and loaded into slide trays. Instructions (similar to those at the left) are read to individual observers or to groups of observers. Subjects are given no other information prior to judging the slides. For example, they are not told the number or nature of the areas to be represented.

Slides are then presented one at a time and each observer records a judgment from the ten-point scenic beauty scale for each slide. For keypunching and computational reasons, a one-digit, ten-point scale (0-9) is preferable, thus all ratings are coded on a 0-9 scale for inputing to our computer program. An example of a rating scale appears on the sample scoring sheet.

We have experimented with numerous rating scales and have found that no significant gain in discrimination is obtained by scales with larger ranges. Scales with a range of 7-10 rating categories have generally been found to be effective for this type of judgment task. Sometimes, there may be some advantage to using scales with more restricted ranges (e.g., 5-7 rating points) if, for example, only a few slide samples

Sample Rating Scale Instructions

"I am going to read some standardized instructions.

Today, more than ever, prudent management of wildlands such as our National Forests is very important. Many wildland researchers are conducting investigations on the effects of alternative vegetative management on water, forage, and timber yields; others are investigating influences on recreational use. In our own research, we are attempting to determine the public's esthetic or scenic perception of such management alternatives, and we greatly appreciate your time in this effort.

We are going to show you, one at a time, some color slides of several wildland areas. Each scene represents a larger area. We ask you to think about the area in which the slide was taken rather than about the individual slide itself.

The first slides will be shown very quickly, just to give you an idea of the range of areas you will be judging. Try to imagine how you would rate these slides, using the "rating response scale" on the top of your scoring sheet. Note that the scale ranges from zero, meaning you judge the area to be very low in scenic quality, to nine, indicating very high scenic quality.

Then, after the initial slides, I will announce that you are to begin rating the next set of slides. You should assign one rating number from zero to nine to each slide. Your rating should indicate your judgment of the scenic beauty represented by the slide. Please use the full range of numbers if you possibly can and please respond to each slide.

Are there any questions before we start?"

or observers are available for each area. Within reasonable limits, SBE values derived from scales with different ranges will not vary appreciably. In fact, research has shown (e.g., see Swetts 1973, Egan 1958, Green & Swetts 1966) that with appropriate (SBE-type) transformations, indices based upon paired-comparisons (forced choice), rankings, or ratings using scales ranging from 4 to 37 points will be comparable.

We have also investigated less "open" scales wherein each point is defined (e.g., 4 = "reasonably certain — I like it"). We found no significant differences between responses to such a "cluttered" scale and those to a more "open" scale. We opted for the above scale, with only the end points defined, because it was easy to describe to the subjects and they seemed to "know how to use it."

Slides are exposed long enough to provide viewers sufficient time to view the slide, make and record a judgment, and prepare for the next slide. In several experiments with different exposure times, we have found that 5 seconds is probably the lower time limit and that 8 seconds is ample under most circumstances. Longer exposures seem to be uncomfortably long; observers tend to get bored or tired. The maximum number of slides that can be shown before judgments become unreliable depends on subject motivation. At an exposure time of 5 to 8 seconds, 100 slides is pressing the upper limits for all but the most motivated (such as college students receiving research participation credit). The decision of how many slides to use is further constrained by (1) the number of areas to be represented, and (2) the number of slides necessary to adequately represent each area.

Evaluating Judgments

After the slide judgment session, score sheets are collected and the ratings key-punched. For analysis we have been using a computer program that has evolved from an earlier version (Wheeler et al. 1971) developed for the Air Force as part of an image-evaluation research project. The present program provides numerous data summaries and computations in addition to SBEs.

Rating Response Scale

Low					Scenic Beauty					High
1	2	3	4	5	6	7	8	9	10

Responses

1. ___	21. ___	41. ___	61. ___	81. ___
2. ___	22. ___	42. ___	62. ___	82. ___
3. ___	23. ___	43. ___	63. ___	83. ___
4. ___	24. ___	44. ___	64. ___	84. ___
5. ___	25. ___	45. ___	65. ___	85. ___
6. ___	26. ___	46. ___	66. ___	86. ___
7. ___	27. ___	47. ___	67. ___	87. ___
8. ___	28. ___	48. ___		
9. ___	29. ___			
10. ___				

The SBE Computer Program

The SBE computer program (SBE) has evolved through several stages of development. The present version is designed so that persons relatively unfamiliar with computers can use it. Outputs have been designed to be easily understood. SBE is capable of making numerous statistical computations that should be of particular interest to researchers. Much of the detailed statistical material is optional, however, and must be requested by the user.

The non-optional outputs include identification pages, mean and median ratings and standard deviations of ratings, frequency, cumulative frequency, cumulative probability, and z-score summary tables, SBEs, and a summary page.

Optional outputs include detailed statistical information regarding frequencies, probabilities, and z scores. In addition to the SBE index (defined as 100 times the d_m index), d' and d_s scores may be requested.

Raw data may be printed and the user may request an analysis of the proportion of ratings above the midpoint of the rating scale. Punched output, for further analysis, is also an option.

Up to 18 areas (landscapes) may be handled in a given experiment, and essentially any number of experiments can be included in any one computer run. Users may select response scales other than the 10-point scale we have utilized. Except for practical considerations, there is no upper limit to the number of observers, and areas may be represented by any number of slides as long as all areas are represented by the same number of slides.

Importantly, SBE permits the user to select either a "by-observer" or a "by-slide" analysis. The overall area SBEs are then averages of the individual slide SBEs or the individual observers SBEs.

User information and the SBE listing require considerable space. We have, therefore, made the "SBE Computer Program" an unattached Appendix to this Paper. Copies may be obtained from:

Individual SBEs may be calculated for each observer based upon his judgments of the several slides representing each area (referred to as "by-observer" analysis). Alternatively, SBEs may be calculated for each slide from an area based upon the judgments made by the different observers (referred to as "by-slide" analysis). The SBE for a given area or landscape is the average of these individual SBEs, either "by-slide" or "by-observer." We have found only minor differences in SBEs determined by the two methods for a given area.

```
                                        ** SUMMARY PAGE **


                          TEST RUN NO. 148 ----- UNIVERSITY UNDERGRADUATES (4/18/74)




TI_A STUDY HAS:
      34 OBSERVERS
      14 SITE SAMPLES PER CONDITION
       8 RANGE LEVELS
       5 CONDITIONS



                                   SCENIC BEAUTY ESTIMATES

                                           *SBE*

     CONDITION 1::      0        WATERSHED 18A
     CONDITION 2::     28        DBSL CANYON LAKE AREA
     CONDITION 3::    -61        WEASER'S LAST STAND
     CONDITION 4::     82        BROWN GULCH (DBST SLIDE)
     CONDITION 5::    -18        CARDER SALE
```

TESTING THE SBE METHOD

The following experiments were conducted primarily to test the reliability and validity of the SBE technique. Each of the experiments followed the general procedure outlined in the preceding section. The major differences between experiments were the observers sampled and the slide samples that represented the evaluated areas. All experiments employed the general statistical models shown below.

General ANOVA Models

RATINGS

Source	df	F
Area	$(a - 1)$	MSA/MSAxO
Observers	$(n - 1)$	
Area x observers	$(a - 1)(n - 1)$	
Slides/areas	$a(s - 1)$	
Observers x slides/areas	$a(n - 1)(s - 1)$	
Total	$a(n)(s) - 1$	

SBEs

Source	df	F
Area	$(a - 1)$	MSA/MSAxO
Observers	$(n - 1)$	
Areas x observers	$(a - 1)(n - 1)$	
Total	$a(n) - 1$	

Note: Degrees of freedom for areas (a) are one less in the SBE than in the ratings analysis because of the deletion of the standard (base) area.

Reliability

The reliability of a scenic beauty measurement method can be assessed in terms of its ability to produce the same measurement (scenic beauty index) on separate applications to (a) the same landscapes, and (b) the same observer population. The SBE Method is a complete assessment system; that is, all phases from the photographic representation of

Outline and Results of Reliability Experiments

1971 Samples

Experiment No.	No. Observers	No. Slides Per Area	Description of Slide Sample	Description of Rating Scale	ANOVA—Ratings (Area Main Effect)		ANOVA—SBEs (Area Main Effect)	
					df	F	df	F
I	30	25	Random Walk (1971)	10-pt Like-dislike	(5/145)	62.65	(4/116)	97.43
II	27	25	Same slides as 1971 new presentation order	10-pt Like-dislike	(5/130)	38.93	(4/104)	55.54
III	23	25	Random walk 1971, new slide sample (with replacement)	10-pt Scenic Quality Scale	(5/110)	40.04	(4/88)	53.79
IV	23	15	Random walk 1971, third sample (with replacement)	10-pt Scenic Quality Scale	(5/110)	51.67	(4/88)	73.18

1972 Samples

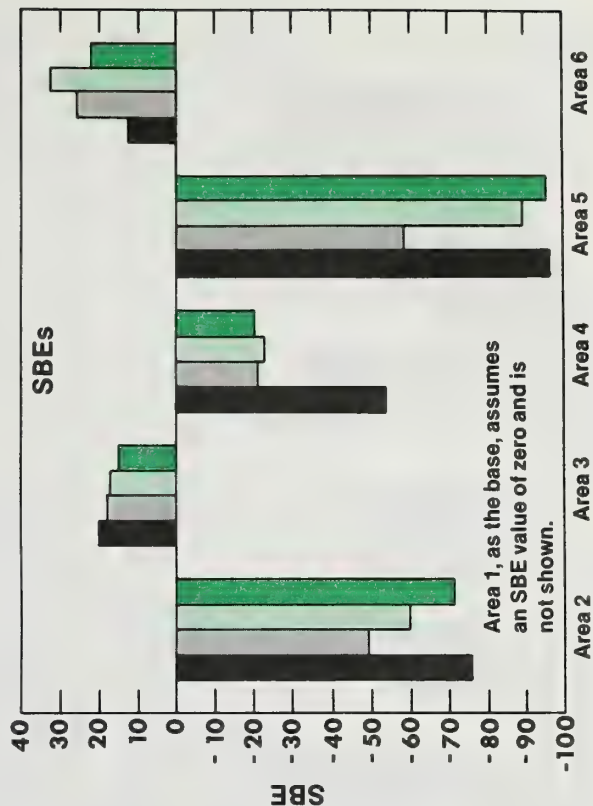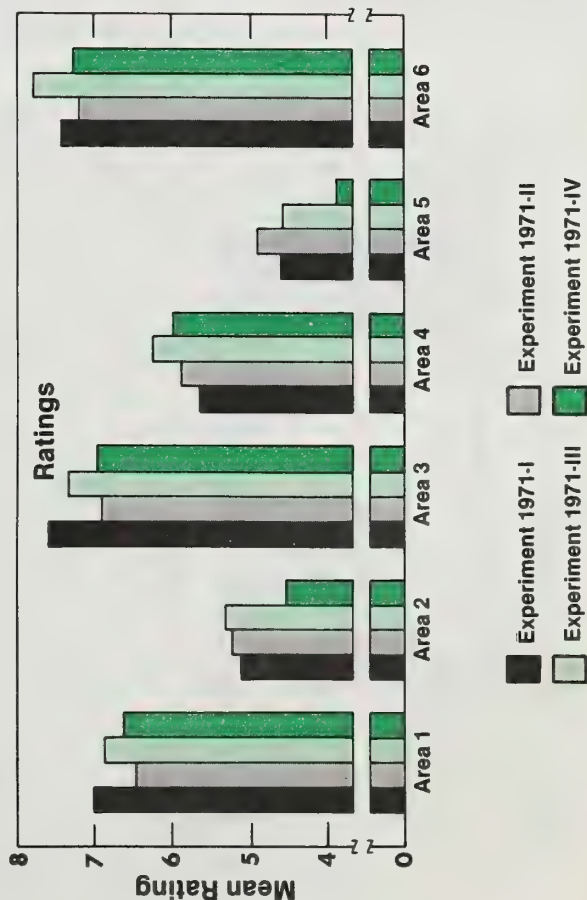| Experiment No.	No. Observers	No. Slides Per Area	Description of Slide Sample	Description of Rating Scale	ANOVA—Ratings (Area Main Effect)		ANOVA—SBEs (Area Main Effect)	
					df	F	df	F
I	40	25	Random Walk (1972-I)	10-pt Scenic Beauty Scale	(5/195)	31.01	(4/156)	36.38
II	16	25	Random Walk (1972-II) (entirely new sample)	10-pt Scenic Beauty Scale	(5/75)	13.13	(4/60)	14.84
III	26	25	Same as Experiment II	10-pt Scenic Beauty Scale	(5/125)	40.97	(4/100)	47.71
IV	33	25	Same as Experiment II	10-pt Scenic Beauty Scale	(5/160)	75.07	(4/128)	81.36

Note: All experiments used different independently sampled observer groups. The slide sample for the 1971 experiments was photographed in October 1971. The 1972 samples were taken in June (1972-I) and July (1972-II). In Experiment IV, 1972, Area 3 was replaced by a forest fire site and Area 5 was replaced by a park-like stand (Last Chance Mine). All values are significant at the .01 level.

the landscape to the selection of appropriate observers, to the final derivation of SBE indices, are integral components of the measurements produced. Thus, it would not be adequate to test the reliability of the method by showing the same slides to the same observers on separate occasions (sometimes referred to as test-retest reliability). All of our reliability tests involved completely independent samples of observers and, in most instances, independent samples of slides to represent the landscape areas being assessed. This is the most appropriate and stringent test of reliability.

Reliability was specifically tested in eight separate experiments generally involving the same six study areas. The six areas were always sampled via the stratified random procedure described earlier. The specific samples, one set taken in 1971 and two separate sets taken in 1972, are described in the table on page 30.

Results of the eight reliability experiments — four experiments based on 1971 samples and four based on 1972 samples — are presented in the figures. The left panel in each figure presents the mean ratings,

SBEs

Area 1, as the base, assumes an SBE value of zero and is not shown.

1971 Sample

Ratings

Experiment 1971-I
Experiment 1971-II
Experiment 1971-III
Experiment 1971-IV

from the 10-point scenic beauty scale, assigned to each of the six landscapes represented. The right panel presents the SBEs.

Because observers for all experiments were from the same population (college students), differences in judgmental criteria among the sample groups were not expected to be large. The very small differences between groups in their ratings of each individual area confirm this expectation. Thus, the SBE analysis would not be expected to alter the pattern of group-to-group comparisons.

By eliminating the effects of differential criteria between observers within the samples, however, SBEs do provide a more precise measure of the differences in scenic beauty among the six areas. Further, the SBEs obtained for each landscape area are independent of the particular rating scale used and the arbitrary judgment criteria adopted by these observers.

While all four experiments based on the 1971 samples produced reliable differences in the judged scenic beauty of the represented landscapes (in other words, the method is sensitive), the primary concern is the extent to which the results of separate applications of the SBE method are consistent. In spite of the fact that the four experiments differed in several ways from each other, both the pattern of results and the actual values found for each landscape are remarkably uniform.
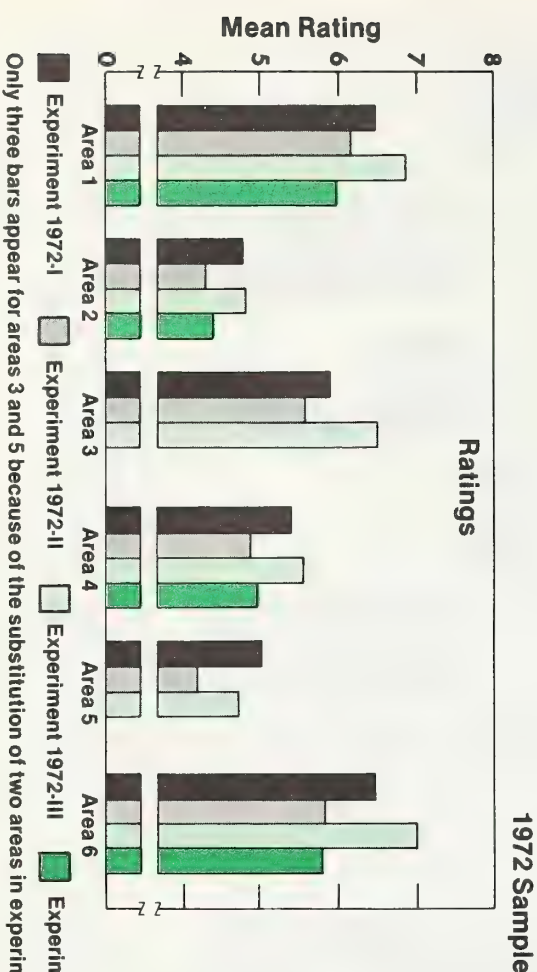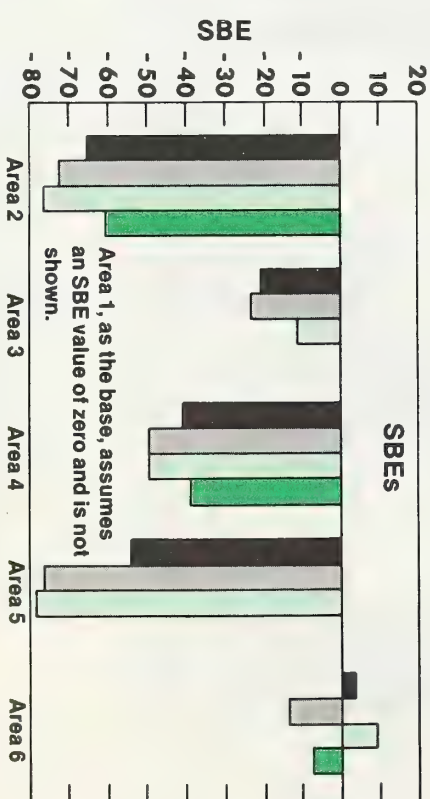
Reliability Experiments

Eight experiments directed toward testing the reliability of the SBE Method were conducted. Six landscape areas were sampled in 1971 and resampled in 1972:

Area 1 — Milk Ranch Point
Area 2 — Beaver Creek Experimental Watershed ("C" in other experiments)

Area 3 — Beaver Creek Experimental Watershed 9
Area 4 — Beaver Creek Experimental Watershed 14 ("B" in other experiments)

Area 5 — Beaver Creek Experimental Watershed 12
Area 6 — Mahan Park

A forest fire site (the Rattle Burn) and an essentially virgin stand (a patented mine claim, Last Chance Mine) were also sampled in 1972 and included in one reliability experiment. All of these areas lie on the Coconino National Forest, Arizona, and are described in detail in Boster and Daniel, 1972 and Daniel, et al. 1973. Independent groups of observers were sampled for each experiment.

Details of the experiments, including the statistical results for both ratings and SBE analyses, are presented in the "Outline" table (page 30). Mean SBE values obtained in the eight reliability experiments are presented in the table to the right.

Ratings

1972 Sample

Mean Rating

Areas 1–6

SBE

SBEs

Area 1, as the base, assumes an SBE value of zero and is not shown.

■ Experiment 1972-I   ▨ Experiment 1972-II   □ Experiment 1972-III   □ Experiment 1972-IV

Only three bars appear for areas 3 and 5 because of the substitution of two areas in experiment 1972-IV.

Mean SBEs for Reliability Experiments

	1971 Samples						1972 Samples					
	Area						Area					
Experiment	1	2	3	4	5	6	1	2	3	4	5	6
I	0	−85	21	−61	−108	12	0	−70	−21	−42	−57	4
II	0	−54	20	−25	−65	27	0	−79	−26	−52	−83	−15
III	0	−67	18	−25	−99	35	0	−83	−12	−52	−85	9
IV	0	−82	16	−24	−108	22	0	−68	Mine	−43	Fire	−8

ANOVA Summaries of the Above

1971 Experiments

Source	df	MS	F
Experiments (E)	3	9,638.29	0.90
Observers/Experiments (O/E)	88	10,670.26	—
Areas (A)	4	252,034.01	233.71 p<.01
Areas X Experiments (AxE)	12	2,857.54	2.65 p<.01
Areas X Observers/Experiments (AxO/E)	352	1,078.39	—
Total	459		

1972 Experiments

Source	df	MS	F
Experiments (E)	2	7,860.18	0.56
Observers/Experiments (O/E)	45	13,853.48	—
Areas (A)	4	56,842.75	56.71 p<.01
Areas X Experiments (AxE)	8	1,855.90	1.85
Areas X Observers/Experiments (AxO/E)	180	1,002.30	—
Total	239		

Note: To avoid the inherent problems of an unequal-N ANOVA, the number of observers in each experiment was randomly reduced to the minimum that would equalize Ns (N = 23 and N = 16 for the 1971 and 1972 experiments, respectively). Also, the fourth experiment was excluded from the 1972 analysis because of the substitution of the "fire" and "mine" areas for two of the six original areas. While the areas X experiments interaction does achieve statistical significance in the 1971 analysis, this source accounts for less than one percent of the total variance (see Hays 1963, p. 406–407). Thus, it is safe to conclude that differences between experiments are very minor.

While the general pattern of results for the 1972 experiments is quite similar to that obtained in 1971, there are several differences. The rankings of the areas are consistent with the 1971 samples, but the ratings of all areas are generally lower and the magnitude of differences between areas is less. This may reflect the fact that the 1971 slides were taken early in the fall (after the rainy season) and both 1972 samples were taken in the summer (prior to the rainy season). Further, specific sub-areas of the base area (Area 1) sampled in 1971 were not resampled in 1972 because a timber harvest was then underway in these areas. For these reasons, detailed comparisons between the 1971 and 1972 samples are probably not warranted.

Although the four experiments based on the 1972 samples differed in many respects, the evaluations of the areas — both in terms of average ratings and SBE indices — were again very similar. The figures show that differences between areas were large and generally consistent from experiment to experiment. Further, the differences from one experiment to another for each area were relatively small. In particular, a comparison of Experiments I and II reveals very consistent results for two completely independent random-walk samples of the same landscapes.

The results of the eight experiments briefly reviewed above indicate that the SBE Method does provide reliable indices. SBE values for specific areas remain essentially the same regardless of changes in the particular slides sampled from the landscape and/or changes in groups of observers sampled from the same population. At the same time, the method provides consistent, quantitative indices of differences in scenic beauty between several forest landscapes.

Validity

Even though a measurement method yields consistent indices from one application to another, there is no guarantee that these indices are related to the quality that the method purports to measure. Determining validity is more difficult and necessarily less direct than establishing reliability.

33

One of our first validity concerns was to determine if the randomly taken color slides adequately represented the actual landscapes being assessed. Because the SBE indices are presumed to apply to the landscapes (not just the color slides), we felt it was important to confirm empirically that slides accurately represent the landscapes. This aspect of our validity testing has been reported in our previously cited publications and a related study is also presented later in this paper. In every case, we have found that color slides, sampled by our random-walk procedures, yield SBE values that are essentially identical to those obtained from observers who make judgments while in the actual areas. The close relationships between slide and onsite SBEs ($r = .98$ and $r = .97$ in two separate tests) convinces us that SBEs determined from color slides provide an accurate estimate of scenic beauty judgments that would be offered if observers were judging the actual landscapes on-site.

In an effort to determine whether the SBE actually measures what it purports to measure, we have compared SBEs against other indicators of scenic beauty. The results of a paired-comparison experiment in which observers chose the most scenically attractive member of successive pairs of landscape panoramas, were compared with SBE values for the same landscapes represented by random walk, color slides. The areas were rank ordered identically by the two methods.

In another test, the stability of the SBE index was evaluated by substantially altering the context in which scenic beauty judgments were made. (This experiment was described earlier as 1972-IV.) Two very different areas — a forest fire site and a park-like mature ponderosa pine stand — were substituted for two of the original six areas. In spite of this rather dramatic change in the context in which the other four areas were judged, the SBEs for the four original areas were essentially unchanged.

Determining the validity of any measuring instrument is a particularly complex problem. Only repeated demonstrations of applicability and

Validity Experiments

Evaluating the validity of the SBE method was approached in several ways. First, the accuracy of the random-walk, color slide representation of landscapes was determined. Second, where other measures of the scenic beauty of the same landscapes were available, SBE values were compared to the other measures. Finally, the context in which measurements were taken was altered by introducing different landscapes in the set being evaluated to determine the stability of the SBE measure.

1: Slide representations: Two experiments compared SBEs based on color slide judgments with SBEs based on judgments made on-site. The relationship was found to be nearly perfect in both experiments, $r = .97$ in the experiment reported in Boster and Daniel (1972) and $r = .98$ in the public-sample validity test reported later in this paper. Both correlation coefficients are highly significant (1% level).

2: Four of the areas that have been evaluated using the SBE method were also compared in an experiment conducted by two colleagues. They collected 180-degree panoramic color-photograph representations of Beaver Creek Experimental Watersheds (our Areas 2, 3, 4, & 5). All possible pairs of panoramas were presented to observers who were required to choose, from each pair, the one they liked best from a scenic quality standpoint. The resulting percent choice data are tabled below along with the SBE values that we obtained for the same areas:

Area	Percent Choice	Rank Based on Percent Choice	SBE	Rank Based on SBE
2	18	3	-66	3
3	29	1	2	1
4	24	2	-37	2
5	5	4	-78	4

The percent-choice data and the SBE measures rank the areas identically. Thus, independently applied methods for evaluating the scenic beauty of these landscapes are in agreement, lending support to the contention that the same underlying property (scenic beauty) is being measured by both methods.

3: The final validity-related experiment was described in the previous section as 1972-IV, and its results are presented in that section. The

utility can truly "validate" a method. While no one of the experiments presented above is sufficient, by itself, to confirm the validity of the SBE method, together they provide strong support for the contention that the SBE is a valid index of the perceived scenic beauty of forest landscapes.

concern here was that the "context" in which evaluations are made should not unduly influence the SBE values obtained. Substituting an area burned by a forest fire and a park-like ponderosa pine stand for a strip cut area and a clear cut area (Areas 3 and 5, respectively) did not change the SBE values for the four areas that were the same in all four experiments reported (Experiment 1972-I-IV). While there are surely limits to the stability of the SBE measure (introducing pictures of the Grand Canyon, for example, might destabilize the SBEs) the SBE measure is stable under reasonable context changes.

USER, INTEREST, AND PROFESSIONAL GROUP TESTS

Development of the SBE Method required that large numbers of subjects participate in numerous experiments designed to test reliability, validity, and utility. For purposes of developing, testing, and refining the method, college students served very well. College students are, of course, an important interest group in themselves, and have been shown to be representative of the total population in many respects. It is imperative, however, that management decisions for public lands reflect the needs and desires of as broad a sample of the general public as possible. Assessment of the scenic preferences of the public is especially important for evaluating land management alternatives, because the scenic impact of land use decisions is probably the most immediate and widespread of all public consequences of management actions.

Assessment of the preferences of the "general public" presents many problems. An undifferentiated "average" of public opinion may not be a useful or meaningful guide to land management policy. It would imply that all public preferences should be weighted equally, and ignores the fact that particular actions in certain areas may have a great deal of impact on one segment of society, but have very little (or even the opposite) effect on other segments.

As a result, public land managers have generally reacted more to specific interest and user groups than to the "general public". As managers are well aware, however, even these relatively narrow interests are often only represented by phone calls, letters, or speeches from individuals who may not reflect the preferences of the group as a whole. More accurate objective assessment of the scenic preferences of relevant public groups is essential for effective public land management.
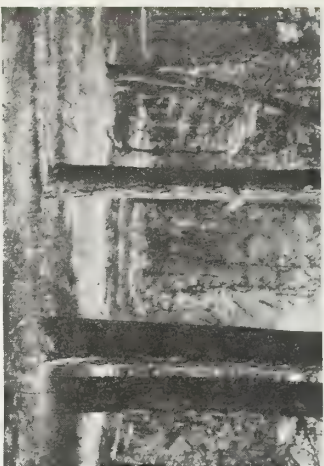
The following experiments tested the applicability of the SBE method for assessing the scenic preferences of a number of user, interest, and professional groups. This phase of our research was scheduled only after we had satisfied ourselves as to the reliability and validity of the technique. Prior to approaching any groups for slide-sample tests, a
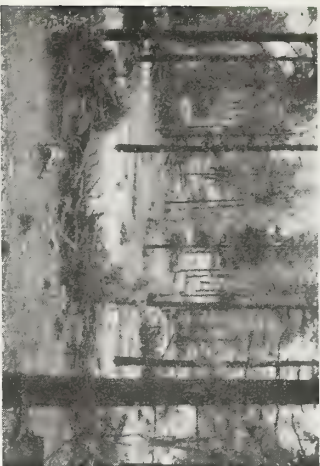
Study Area Descriptions

A

WOODS CANYON

Large (10,700 acres), moderately heterogenous area last logged about 1950. Average basal area approximately 75 sq. ft./acre, but tree density is variable.

B

BEAVER CREEK EXPERIMENTAL WATERSHED 14

Irregular clearcut strips averaging 60 feet wide. Intervening (leave) strips thinned to 80 sq. ft./acre basal area. Slash burned. Also known as the "golf-course" treatment. Logged 1970–1, 1350 acres (for more detailed information, see Turner and Larson 1974).

C

BEAVER CREEK EXPERIMENTAL WATERSHED 17

Thinned from approximately 100 sq. ft./acre basal area. Slash piled into (high) wind rows. Logged 1969, 299 acres (for more details, see Turner and Larson 1974).

D

WILD BILL EXPERIMENTAL UNIT 5

Fairly homogenous area, thinned to 60 sq. ft./acre basal area. Substantial grass cover. Logged 1964, 164 acres (for more details, see Pearson and Jameson 1967).
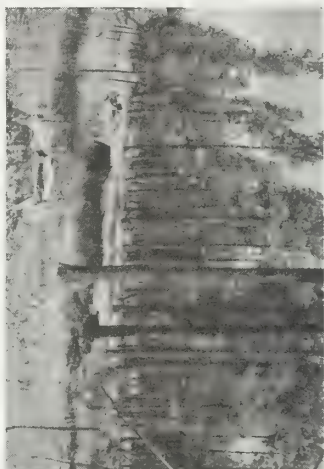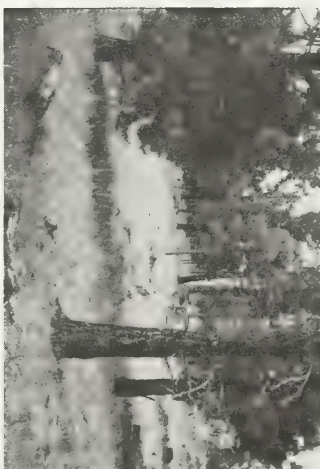
E

HORSELAKE SALE

Conventional timber sale (Shelterwood silvicultural system), 6500 acres, logged 1967, 3.3 MBF/acre harvested, approximately 8 MBF/acre remaining, slash lopped, no pulp sale, no burning.

F

FRED HAUGHT RIDGE SALE

Intensive silvicultural management, 5000 acres, sawtimber harvest 1970 (original virgin cut 1950), pulping sale 1972, precommercial thinning, slash lopped to less than three feet.

All areas are on the Coconino National Forest (Arizona). Although some oak is occasionally evident, the areas are essentially ponderosa pine forests. Areas E and F have extensive fuel break systems. However, all photo samples were taken outside break zones, where there is considerable downed wood. Therefore, the scenic judgments offered for these two areas are not likely to be indicative of the areas as a whole (i.e. with fuelbreaks).

38

complete field test, including bussing observers to the actual forest sites, was conducted with a sample public group. In addition to testing and further refining the SBE methodology, we sought to determine what differences may exist among public groups in the perception of landscape scenic beauty.

Groups

A total of 26 public groups provided evaluations of landscape areas representing several realistic forest management alternatives. These groups varied in many respects. Some represented rather localized concerns (such as the Flagstaff Beauty Commission). Others represented State-wide or regional interests (such as the Arizona Cattlegrowers Association and the Southern Arizona Environmental Council), while other groups were sampled from a much larger geographic area (Regional Landscape Architects, U.S. Forest Service). The number of individuals within each group ranged from 7 to 114. The representativeness of the specific sample obtained for each group is undetermined, so any implications of the results for the group as a whole should be approached with caution.

Areas

Our earlier work was done on areas that differed markedly from each other in terms of vegetative management. Several of the areas were experimental watersheds not intended as operational models. The areas shown to the interest and user groups, however, comprise a cross section of management options in Arizona's ponderosa pine National Forests. Areas were selected with the help of Forest Service personnel at the Regional Office, Southwest Region.

All six areas are on the Coconino National Forest in Arizona. They are described briefly in the adjacent figure. A "representative" picture of each is also included. Because experience has taught us to avoid generalized descriptive names for complex silvicultural prescriptions, we therefore refer to these areas by letter.

These areas were photo sampled by the random-walk procedure described earlier. Fifteen slides were randomly selected for each area

User, Interest, and Professional Groups (N)

Data Summary

| | Ratings (Ranks) | | | | | | Area Main Effect |
	A	B	C	D	E	F	F(5,84)
RANGE INTERESTS							
1 Ariz. Cattlegrowers Assoc. (40)	4.66 (4)	5.51 (2)	5.33 (3)	5.69 (1)	4.33 (5)	4.01 (6)	7.19**
2 Ariz. Cowbells Assoc. (16)	4.57 (4)	5.20 (2)	4.85 (3)	5.50 (1)	4.17 (5)	3.72 (6)	5.06*
3 Ariz. Section, Soc. Range Management (15)	5.56 (3)	5.99 (2)	5.34 (3)	6.75 (1)	5.29 (4)	5.01 (5)	9.78***
4 Tonto National Forest, Supervisor & Staff (13)	4.55 (3)	5.36 (2)	4.35 (4)	6.45 (1)	4.13 (5)	4.09 (6)	10.51***
FORESTERS							
5 Coconino National Forest, Supervisor & Staff (17)	5.51 (2)	5.33 (3)	4.11 (6)	6.10 (1)	4.82 (4)	4.18 (5)	6.87**
6 Professional Staff, Regional Headquarters (R-3), USFS (23)	5.52 (3)	5.54 (2)	4.11 (6)	6.33 (1)	4.72 (5)	5.43 (4)	11.06***
7 Scientists, Rocky Mtn. For. & Range Exp. Station, Tempe, Ariz. (16)	5.53 (2)	5.38 (3)	4.22 (6)	5.89 (1)	4.83 (4)	4.40 (5)	5.40**
8 Forestry Club, U. of Arizona (11)	5.56 (2)	5.51 (3)	4.72 (6)	5.81 (1)	5.10 (4)	4.91 (5)	4.33**
9 Forestry Upperclassmen, No. Ariz. University (92)	4.86 (3)	5.32 (1)	4.51 (5)	4.70 (4)	4.97 (2)	4.33 (6)	3.31**
ENVIRONMENTALISTS							
10 Tucson Chapter, Sierra Club (16)	5.51 (2)	5.20 (3)	3.80 (6)	5.84 (1)	5.09 (4)	4.59 (5)	7.72**
11 Southern Arizona Environmental Council (21)	5.97 (1)	5.57 (3)	4.15 (6)	5.78 (2)	5.31 (4)	4.86 (5)	7.57**
12 Flagstaff Beauty Commission (7)	5.69 (2)	5.24 (3)	3.87 (6)	5.97 (1)	5.19 (4)	4.68 (5)	8.71**
13 Research Staff, Museum of Northern Arizona (18)	5.93 (1)	5.56 (3)	4.70 (6)	5.81 (2)	5.22 (4)	4.80 (5)	4.35**
14 Southern Arizona Hiking Club (32)	5.35 (1)	5.08 (2)	4.20 (6)	5.01 (3)	4.69 (4)	4.50 (5)	6.47**
FOREST ECONOMISTS AND LANDSCAPE ARCHITECTS							
15 Western Forest Economists (Sub-Group 1) (30)	4.50 (3)	4.52 (2)	3.79 (6)	5.14 (1)	4.26 (5)	4.27 (4)	5.53**
16 Western Forest Economists (Sub-Group 2) (30)	4.71 (2)	4.59 (3)	3.82 (6)	5.25 (1)	4.46 (4.5)	4.46 (4.5)	5.46**
17 Western Forest Economists (Sub-Group 3) (54)	4.71 (2.5)	4.71 (2.5)	4.07 (6)	5.03 (1)	4.63 (4)	4.46 (5)	3.76**
18 Forest Service Landscape Architects (29)	4.39 (3)	4.42 (2)	3.59 (6)	4.97 (1)	4.07 (4)	3.75 (5)	5.47**
STUDENTS							
19 Intro. Psychology Students, U. of Arizona (Group 1) (23)	5.41 (2)	5.07 (4)	4.76 (6)	5.82 (1)	5.19 (3)	5.02 (5)	2.21
20 Intro. Psychology Students, U. of Arizona (Group 2) (23)	5.17 (2)	4.75 (5)	4.57 (6)	5.67 (1)	4.88 (3)	4.85 (4)	3.05*
21 Land Use Planning Class, U. of Arizona (25)	5.91 (2)	5.61 (3)	4.17 (6)	6.19 (1)	5.31 (4)	4.77 (5)	7.97**
22 Watershed Measurement Class, U. of Arizona (22)	5.78 (2)	5.60 (3)	4.41 (6)	5.82 (1)	5.33 (4)	4.79 (5)	7.78**
OTHER GROUPS							
23 Tucson Varmit Callers (30)	5.38 (2.5)	5.43 (1)	4.86 (4)	5.38 (2.5)	4.67 (6)	4.75 (5)	2.40*
24 Catholic Adult Group (Tucson Church) (38)	5.85 (2)	5.63 (3)	4.55 (6)	5.97 (1)	5.44 (4)	4.98 (5)	4.54**
25 Agua Prieta, Sonora, Mexico, High School Students (29)	6.44 (1)	5.77 (4)	5.64 (5)	5.94 (3)	6.06 (2)	5.41 (6)	2.17
26 Catholic Teenage Group (Tucson Church) (10)	6.65 (1)	6.13 (4)	5.31 (6)	6.52 (2)	6.27 (3)	6.04 (5)	2.67*

*p < .05
**p < .01

User, Interest, and Professional Groups (N)

Data Summary	SBEs (Ranks)						Area Main Effect	
	Area						(df,df)	F
	A	B	C	D	E	F		
RANGE INTERESTS								
1 Ariz. Cattlegrowers Assoc. (40)	0(4)	29(2)	15(3)	35(1)	−12(5)	−23(6)	4,156	43.57**
2 Ariz. Cowbells Assoc. (16)	0(4)	20(2)	9(3)	32(1)	−14(5)	−31(6)	4,60	18.11**
3 Ariz. Section, Soc. Range Management (15)	0(4)	15(2)	9(3)	41(1)	−10(5)	−19(6)	4,56	22.40**
4 Tonto National Forest, Supervisor & Staff (13)	0(4)	29(2)	6(3)	65(1)	−14(5)	−16(6)	4,48	20.98**
FORESTERS								
5 Coconino National Forest, Supervisor & Staff (17)	0(2)	−7(3)	−47(5.5)	20(1)	−24(4)	−47(5.5)	4,64	31.85**
6 Professional Staff, Regional Headquarters (R-3), USFS (23)	0(3)	1(2)	−49(6)	27(1)	−3(4)	−28(5)	4,88	51.21**
7 Scientists, Rocky Mtn. For. & Range Exp. Station, Tempe, Ariz. (16)	0(2)	−4(3)	−44(6)	12(1)	−23(4)	−39(5)	4,60	30.84**
8 Forestry Club, U. of Arizona (11)	0(2)	−3(3)	−31(6)	7(1)	−16(4)	−24(5)	4,40	10.28**
9 Forestry Upperclassmen, No. Ariz. University (92)	0(2)	16(1)	−12(5)	−5(4)	−4(3)	−18(6)	4,360	18.64**
ENVIRONMENTALISTS								
10 Tucson Chapter, Sierra Club (16)	0(3)	−9(1)	−57(6)	12(2)	−14(4)	−32(5)	4,60	23.97**
11 Southern Arizona Environmental Council (21)	0(1)	−14(3)	−63(6)	−7(2)	−23(4)	−38(5)	4,80	18.99**
12 Flagstaff Beauty Commission (7)	0(2)	−14(3)	−61(6)	10(1)	−16(4)	−34(5)	4,24	8.93**
13 Research Staff, Museum of Northern Arizona (18)	0(1)	−12(3)	−42(6)	−3(2)	−25(4)	−39(5)	4,68	9.04**
14 Southern Arizona Hiking Club (32)	0(1)	−9(2)	−39(6)	−12(3)	−23(4)	−29(5)	4,124	15.38**
FOREST ECONOMISTS AND LANDSCAPE ARCHITECTS								
15 Western Forest Economists (Sub-Group 1) (30)	0(3)	1(2)	−25(6)	22(1)	−8(4.5)	−8(4.5)	4,116	39.23**
16 Western Forest Economists (Sub-Group 2) (30)	0(2)	−4(3)	−31(6)	19(1)	−9(4.5)	−9(4.5)	4,116	35.05**
17 Western Forest Economists (Sub-Group 3) (54)	0(2)	−1(3)	−22(6)	10(1)	−9(5)	−3(4)	4,212	22.86**
18 Forest Service Landscape Architects (29)	0(3)	3(2)	−26(6)	20(1)	−10(4)	−22(5)	4,112	49.45**
STUDENTS								
19 Intro. Psychology Students, U. of Arizona (Group 1) (23)	0(2)	−12(4)	−22(6)	15(1)	−7(3)	−13(5)	4,88	7.05**
20 Intro. Psychology Students, U. of Arizona (Group 2) (23)	0(2)	−15(5)	−21(6)	16(1)	−11(3)	−12(4)	4,88	5.04**
21 Land Use Planning Class, U. of Arizona (25)	0(2)	−11(3)	−60(6)	9(1)	−21(4)	−39(5)	4,96	37.95**
22 Watershed Measurement Class, U. of Arizona (22)	0(2)	−6(3)	−47(6)	2(1)	−16(4)	−34(5)	4,84	24.86**
OTHER GROUPS								
23 Tucson Varmit Callers (30)	0(1.5)	2(3)	−18(4)	0(1.5)	−25(6)	−21(5)	4,116	7.76**
24 Catholic Adult Group (Tucson Church) (38)	0(2)	−8(3)	−45(6)	4(1)	−14(4)	−29(5)	4,146	27.91**
25 Agua Prieta, Sonora, Mexico, High School Students (29)	0(1)	−24(4)	−27(5)	−18(3)	−13(2)	−36(6)	4,112	4.71**
26 Catholic Teenage Group (Tucson Church) (10)	0(1)	−18(4)	−44(6)	−4(2)	−13(3)	−20(5)	4,36	6.05**

Results

Average SBE values for the six areas for each interest group tested are presented in the histoform graphs. Area A, as the selected standard of comparison, assumes an SBE value of zero for all groups.
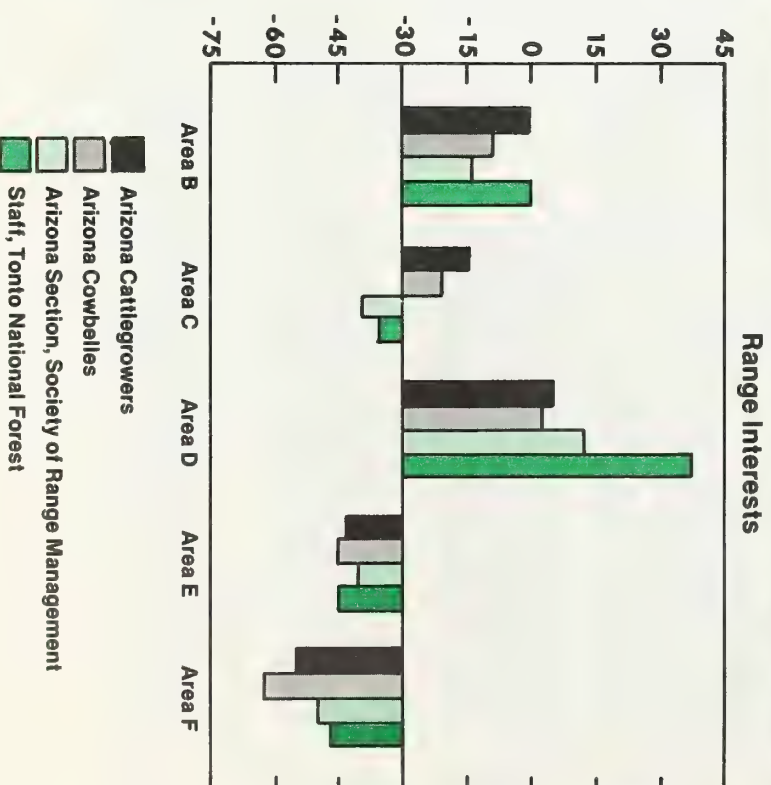
The graphs clearly show that all groups distinguish among the sampled landscapes. Statistical analyses of each individual group's judgments (both ratings and SBEs) confirmed that evaluations differed reliably among landscapes. Further, analysis across the groups indicated statistically significant differences between areas. Thus, there were reliable differences in the way in which groups reacted to the sampled areas.

The principal purpose of these experiments was to compare esthetic perceptions among the various groups sampled. Specific groups have been organized into larger categories for comparison. The assignment of groups to categories was based, for the most part, upon the declared interests and affiliations of the group. Other categorizations are possible, but the ones presented are useful for comparing similarities and differences among the groups represented.
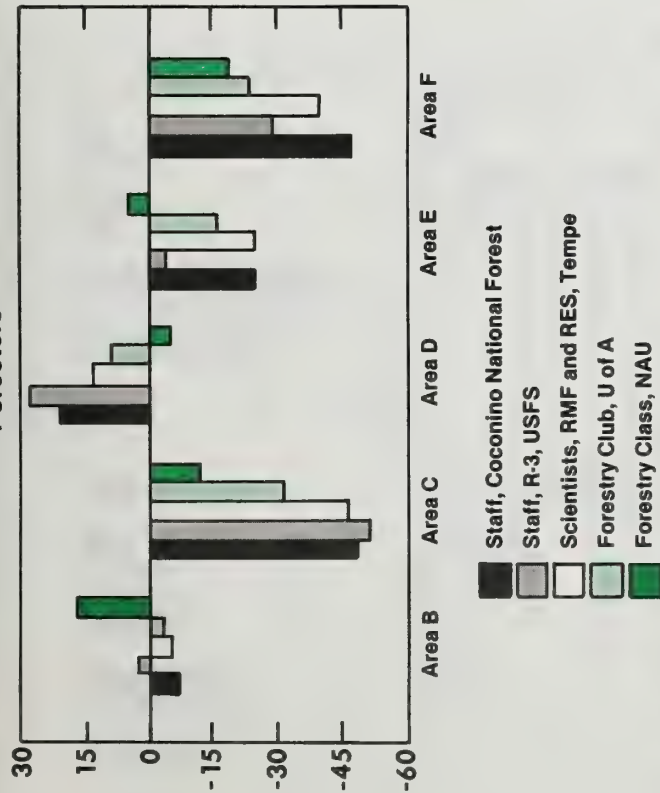
Range Interests

This category includes those groups whose primary interests and concerns with regard to forest management center around cattle growing. The Tonto National Forest staff group was included in this category rather than the forester category because of heavy program emphasis on range management.

The SBE patterns show relatively high agreement between the range groups. All of these groups showed greater preference (relative to Area A) for areas B and D, and judged area C approximately as scenic as area A, the selected standard. These three areas (B, C, and D) generally have the lowest tree densities of the six areas. Apparently (and understandably), cattle growers and the other range interests sampled perceived less dense, and more "grassy", forest landscapes as being higher in scenic beauty than denser stands.

Range Interests

Range Interests

■ Arizona Cattlegrowers
□ Arizona Cowbelles
▨ Arizona Section, Society of Range Management
▧ Staff, Tonto National Forest

Foresters

Groups in this category are primarily representative of forestry and silvicultural interests. Three professional groups (Coconino National Forest Supervisor and staff, professional staff at the Regional Office, Southwest Region, U.S. Forest Service, and Forest Service scientists, Rocky Mountain Forest and Range Experiment Station, Tempe, Arizona), the Forestry Club, University of Arizona, and a class of forestry school students were represented. With the exception of the Northern Arizona University (NAU) forestry class, the agreement between groups in this category is quite good. We can only speculate about the reasons for the rather marked differences between the scenic beauty judgments of the NAU forestry students and those of the other forester groups. It is possible that the physical setting in which the slides were shown was not optimal; this was one of the largest groups we have tested at one time. Also, several of the areas (specifically B, C, and D) had been visited by many of the students as part of their studies.

As did most groups tested, the forester groups judged Area D to be the most scenic. They also perceived Areas C and F to be substantially less attractive (both are characterized by large amounts of piled, downed wood and slash). In addition, Area C is sparsely wooded, approximately three-fourths of the trees have been removed to meet an experimental watershed treatment prescription. It is interesting to note that the most preferred area, Area D, represents an intensive silvicultural treatment; it generally has very little downed wood, and has a vigorous ground cover of forbs and grasses.

Environmentalists

This category represents a wide variety of groups, but all share a common, active interest in the well-being of the natural environment. There was generally strong similarity in the perception of the six landscapes among the groups in this category. Area D was generally judged to be more beautiful than Areas B, C, E, or F, but it was judged to be essentially the same as the "base", Area A.

Area A is the closest of the six to a "natural" area. It was last harvested

43

Foresters

Area B   Area C   Area D   Area E   Area F

■ Staff, Coconino National Forest

▨ Staff, R-3, USFS

☐ Scientists, RMF and RES, Tempe

▨ Forestry Club, U of A

▨ Forestry Class, NAU

approximately 25 years ago, and shows minimal evidence of man's hand. Downed wood is mostly in the form of large trees and other natural-appearing litter. Variation in tree sizes and densities is great, ranging from small open stands of large yellow pines to dense thickets of very small-diameter trees. Overall tree density is greater than for any other area.
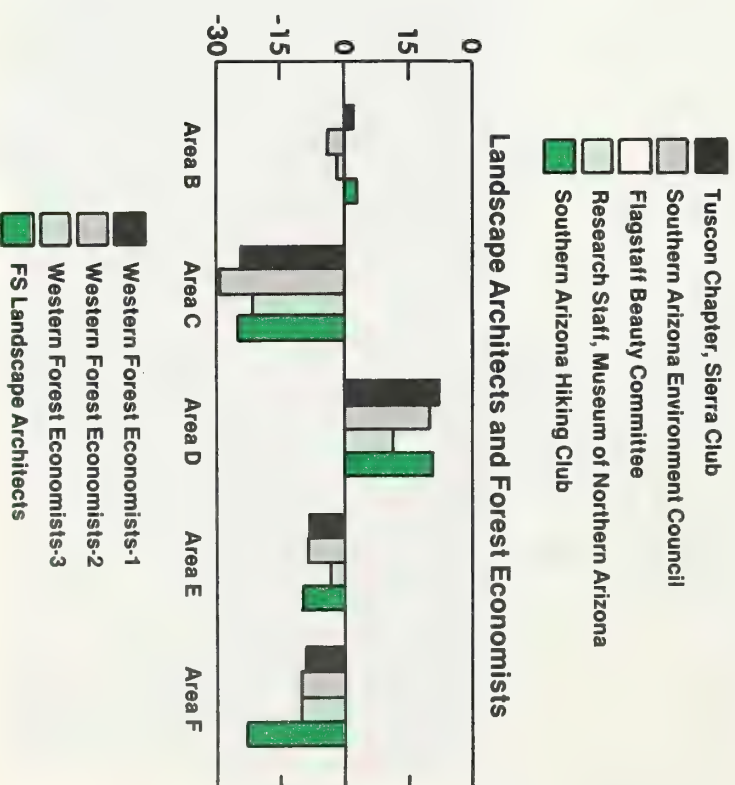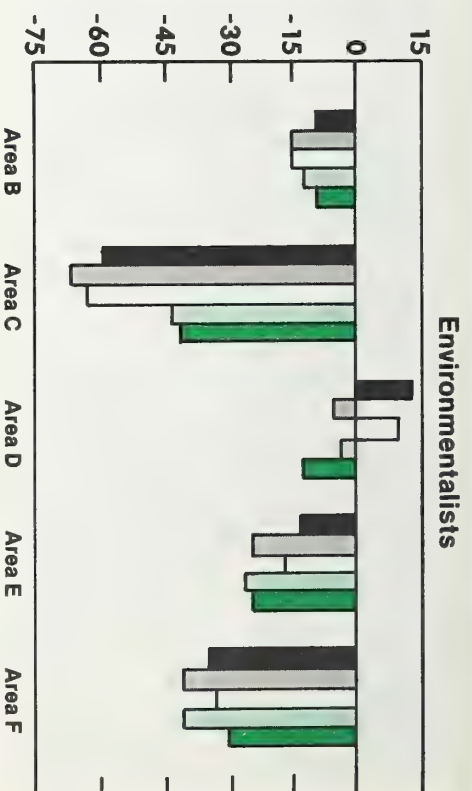
The environmentalist category produced the strongest of the generally negative judgments for Area C. In many ways, this is the most obviously "treated" (unnatural) of all the areas sampled, with very prominent windrowed slash and other evidence of severe thinning. At the same time, however, the intensively managed Area D was perceived to be as scenic as the more "natural" Area A. This would seem to indicate that intensive management is not necessarily detrimental to scenic beauty, even in the eyes of some of the most critical observers.

Landscape Architects and Forest Economists

The pairing of these rather different groups was largely a matter of graphic convenience. Members of both groups are spread over large geographic areas in contrast to the other groups which were comprised of Arizonans. The economists were tested at the annual meeting of the Western Forest Economists, a group that includes economists from private industry, public agencies, and universities; the landscape architects are Forest Service employees stationed in various parts of the country.

The economists were consistent in preferring Area D and disliking Area C, relative to the base area, A. In general, these groups did not distinguish sharply, with regard to scenic beauty, among the landscapes represented. Some indication of the reason for this is provided by the fact that these groups generally rated all of the areas lower than did the other (Arizona) groups. Further, there was some sentiment expressed mostly by individuals from the Northwest, that "all ponderosa pine forests look alike."

The landscape architects' perception of the relative scenic beauty of the ponderosa pine landscapes follows a pattern very similar to that of the forester groups, but is more restricted in range. This restricted range seems to be a result of factors similar to those expressed for the

Environmentalists

Tuscon Chapter, Sierra Club
Southern Arizona Environment Council
Flagstaff Beauty Committee
Research Staff, Museum of Northern Arizona
Southern Arizona Hiking Club

Landscape Architects and Forest Economists

Western Forest Economists-1
Western Forest Economists-2
Western Forest Economists-3
FS Landscape Architects

economist group. The landscape architects produced lower overall ratings than did any other group sampled. In agreement with most of the other groups, they found Area D to be the most beautiful and Area C to be the least beautiful.
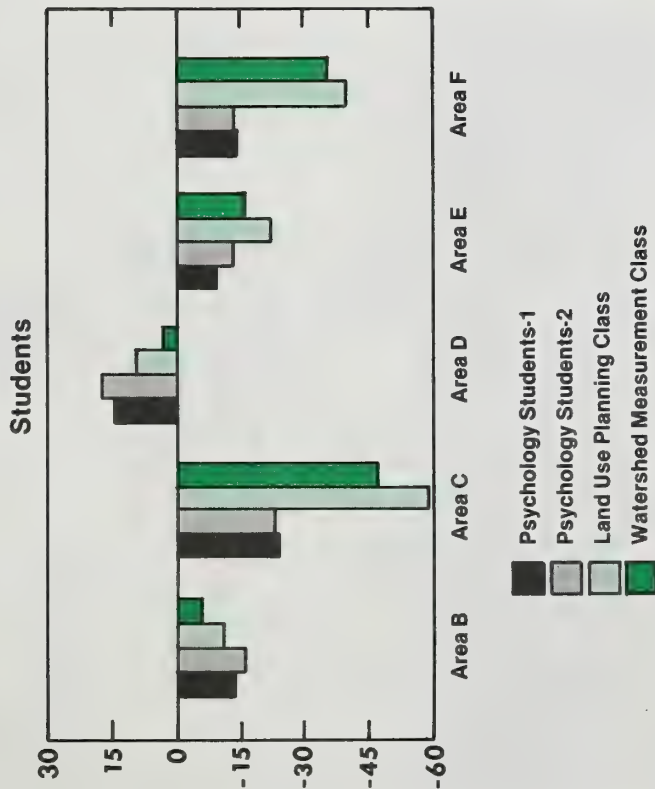
Students

This category was included, in part, because of the large number of student subjects that have been used in experiments concerned with the development of the SBE technique. Also, however, we felt that these groups of mostly young adults represent a large and concerned segment of the general public. Further, it is important to know how students compare with other population sub-groups in order to determine their suitability as "public" stand-ins in any future studies.

The pattern of student SBEs revealed two sub-categories: psychology students and students enrolled in watershed management and land-use planning. The psychology students were mostly freshmen and sophomores, presumably with relatively little knowledge of forests and forest management. The students who were pursuing majors directly related to natural resource planning and management were mostly juniors and seniors. Thus, it is not surprising that they exhibited greater differentiation in their perception of the forest landscapes than the psychology students.

The pattern of results was essentially the same for all of the student groups in spite of the generally poorer differentiation by the psychology students. All agreed that Area D was the most beautiful landscape and that Area C was the least beautiful. In this and other respects, the students' pattern of reactions closely approximates the perception of the foresters. The watershed and land use planning students (many of whom were majoring in forestry or range management) produced results that are essentially identical to the majority of the forester group results.

Other Groups

This category, containing the remainder of the groups sampled, includes a sportsman's group (Tucson Varmit Callers), two groups from

Students

Area B   Area C   Area D   Area E   Area F

Psychology Students-1
Psychology Students-2
Land Use Planning Class
Watershed Measurement Class

a suburban Catholic church, and a group of high school students from Agua Prieta, Sonora, Mexico. For the latter group, the instructions and response scale sheets were in Spanish.

For the most part, the pattern of results for these groups is similar to that of the environmentalists, although the range of responses is smaller. Area D was preferred to all but the base area, A. Area C was, on the average, perceived to be the least beautiful.

The adult Catholic group (in Tucson, Arizona) is of particular interest in that it may offer some indication of the scenic preferences of the "general public". While the group members are certainly not a "representative sample" of the general public, neither do they represent any particular interests with regard to forest management or forest esthetics. Because participants volunteered in response to an announcement describing the nature of the test, however, the actual sample may have been somewhat biased towards people having a higher than average interest in the natural environment.

The scenic beauty evaluations obtained for this group, and generally for their children, grades 7 to 12, are quite similar to those of the forester and environmentalist samples. Area D was preferred over all of the other areas, though only slightly more than Area A. Area C was again the least preferred.

Categories Compared

To facilitate comparisons among the various interests represented in the group samples, SEBs were averaged within the categories described above. Thus, for each area, an average SBE was obtained for range interests, foresters (minus the NAU class), environmentalists, landscape architects, economists, the Catholic adult group, and the psychology students. The SBEs for these categories are graphed in the left panel of the adjacent figure. The right panel presents mean ratings for the same categories.
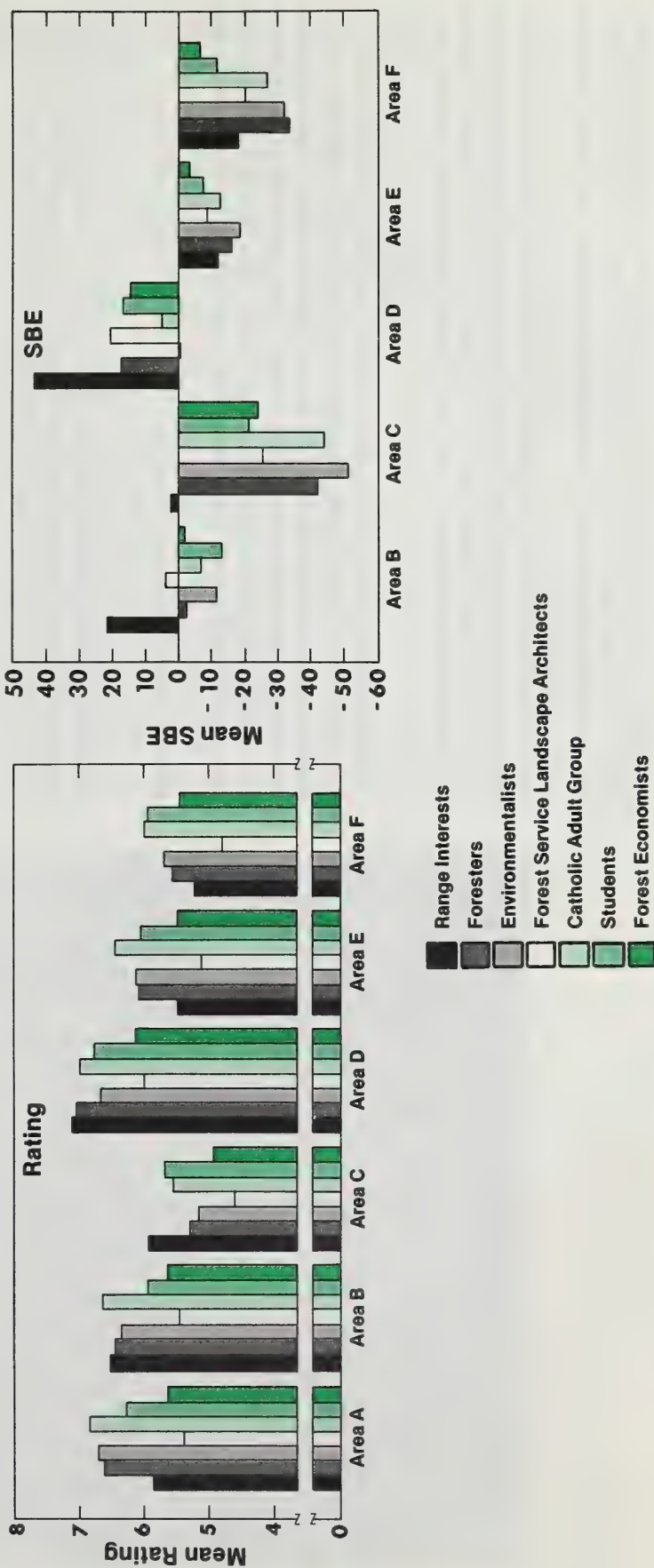
Perhaps the most striking feature of these data is the extent to which there was agreement in the scenic preferences of such diverse interest

Other Groups
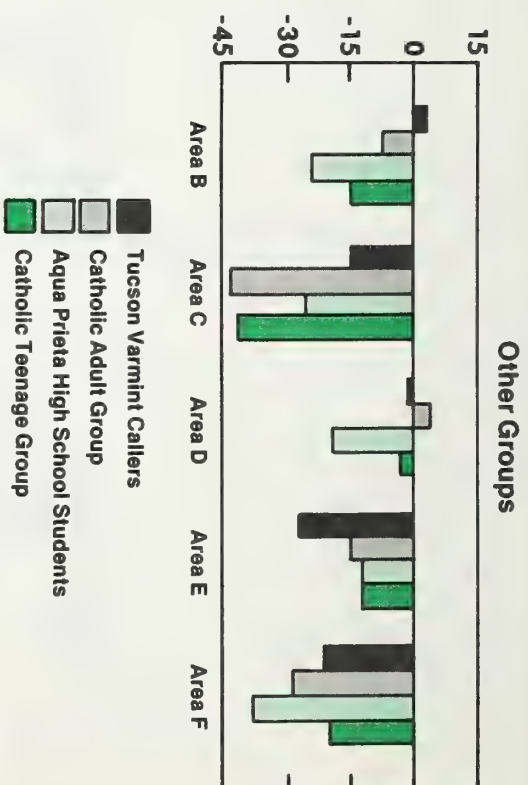
groups. Area D was generally the most preferred, Area C the least. Only the range interests were an exception to this basic pattern.

The mean ratings suggest the essential similarity in the groups' responses, but the pattern is not as easily discerned as in the SBE graph. The effects of varying standards among the groups represented tend to obscure the results. Notice, for example, that the landscape architect group tended to rate all of the areas substantially lower than did any of the other categories shown, suggesting that their judgmental standards were generally higher. In this context, it is interesting to note that the economists — the only other non-Arizona group — were second only to the landscape architects in applying generally low ratings to Arizona ponderosa pine forests.

SBEs and Ratings for Group Categories

Range Interests
Foresters
Environmentalists
Forest Service Landscape Architects
Catholic Adult Group
Students
Forest Economists

While all groups tended to react favorably to Area D, the range interests found this landscape particularly attractive. Areas B, C, and D are relatively "open", and show a generally well developed ground cover, especially grass. Area D was developed as an experimental pasture to study tradeoffs between timber and range production.

The foresters and environmentalists generally perceived the landscapes very similarly. The major differences are that environmental groups responded more negatively to Area C (and to a lesser extent Area B) than did foresters, and foresters responded more positively to Area D than did environmentalists. It is interesting to note that the Catholic adult group represented a compromise between the foresters and the environmentalists in that their SBEs were less negative than the environmental groups for Area C, but lower than the foresters for Area D.

Other similarities and differences among the groups and categories samples may be seen in the SBE graph. The overall picture presented by the SBE analysis, however, is more one of agreement than disagreement.

Public Field Test

The evidence is strong from our own work with student samples (Boster and Daniel 1972) and from similar research by others (Zube et al. 1974, Coughlin and Goldstein 1970) that color photos and slides adequately represent natural landscapes. However, it was felt that an additional on-site validation of the method utilizing the six "new" areas and with a cross section of persons most interested in National Forest land use planning was desirable.

The Forest Supervisor, Tonto National Forest (headquartered in Phoenix), drew up a list of names and sent out a letter of invitation. Invited participants included persons that had shown a strong interest professionally and/or personally in National Forest management and decision making. The roster included five representatives each from the State Wildlife Federation chapter and the Arizona Conservation Council. In addition, there were representatives of the education field (teachers and high school students), Scouting, landscape architecture,

48

timber and cattle industries, Friends of the Earth, Sierra Club, media, campers (the American Camping Association), and water-oriented interests. Remarkably good agreement was found between slide-delivered SBEs and the SBEs for the user, interest, and professional groups previously tested. The following tabulation compares the SBEs of field trip participants who viewed the same slides as did the groups with the average SBEs of all the groups:

Area	Field Test Group	Average of Groups
A	0	0
B	-2	-3
C	-31	-37
D	12	13
E	-20	-16
F	-27	-30

The 34 participants visited each of the six areas that had been photo-sampled and shown to the user, interest, and professional groups. However, one of the areas visited (Area A) was adjacent to the actual, photo-sampled area which is accessible only by foot or 4-wheel-drive vehicle. Before visiting the first area, the participants were randomly divided into two groups: the first half viewed and judged the same slides as had the user, interest, and professional groups; the second half viewed slides taken in the study areas just 2 weeks prior to the field trip. Because one purpose of the trip was to again validate the color-slide representations, we felt the slides to be used in the comparison with on-site judgments should be as recent as possible to avoid any confounding elements, such as seasonally induced vegetative changes.

The more recent slide sample consisted of eight slides — traversing 360 degrees — taken at each of the 18 sites (three sites in each of six areas). Five slides from each set of eight were randomly selected to represent each site. The resulting 90 slides were scrambled and shown to the appropriate participants who assigned esthetic ratings.

All participants were bussed to the three separate sites within each of the six areas sampled. Every effort (including the introduction of a number of "phoney" sites) was made to make the three judgments in

49

Scenic Beauty Field Trip
September 7–8, 1974
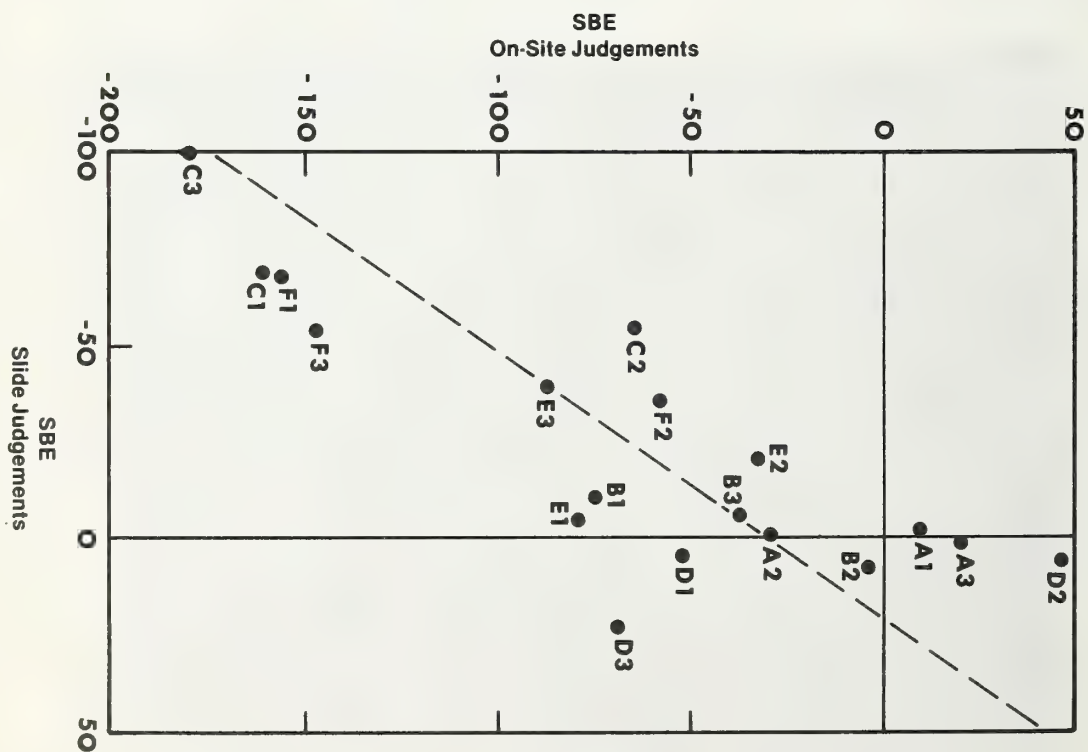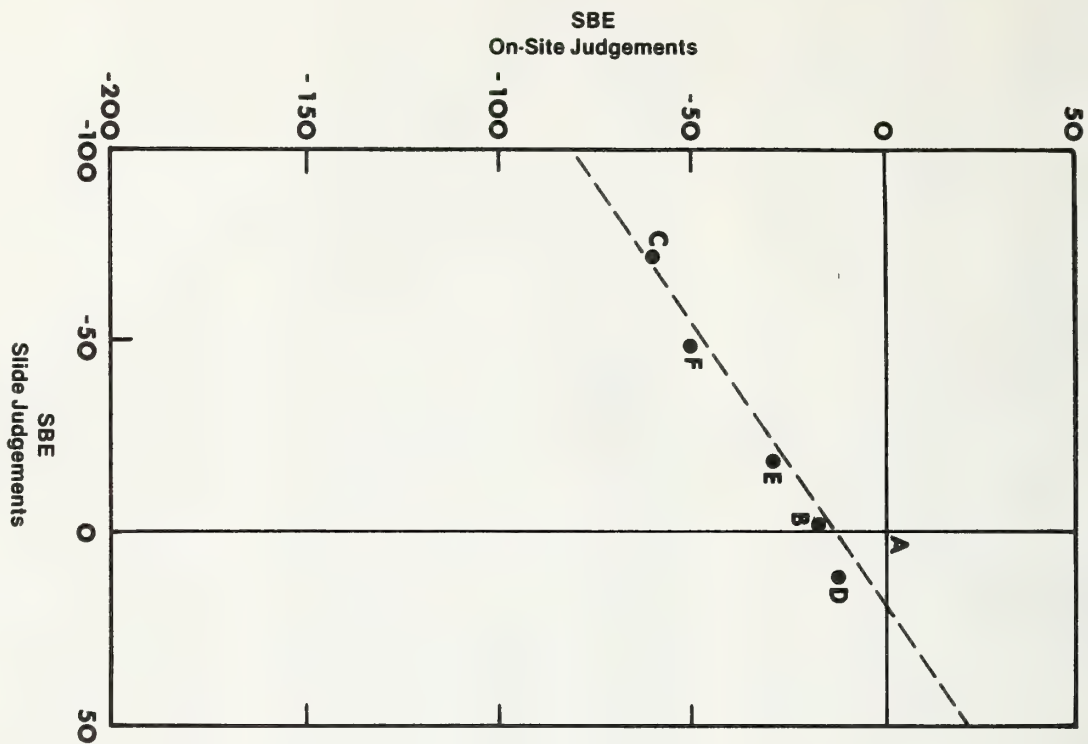
each area as independent as possible. All participants recorded a scenic beauty judgment (10-point scale) for each of the 18 sites visited.

The results can be seen clearly in the paired scatter diagrams which show the slide-derived SBEs plotted against the SBEs obtained from on-site judgments. A perfect agreement between on-site and slide judgments would place all of the points on a 45 degree diagonal line that passes through the origin (0,0) of the diagram. The obtained relationship is slightly off this perfect relationship, primarily because the SBEs derived from on-site judgments are generally lower than SBEs derived from slide judgments. This small, constant difference may have arisen because of the substitution of base areas (area A, explained above) or because all the slide judgments were made prior to any on-site judgments. Regardless, this constant difference would not affect the statistical relationship between the two sets of SBE values. In fact, the correlation coefficients are highly significant statistically.

The results of this particular experiment, coupled with our earlier field test and the work of others, shows conclusively that color slides can adequately represent forest landscapes. The fact that slides do as well as on-site observations makes direct application unnecessary, except perhaps in special cases.
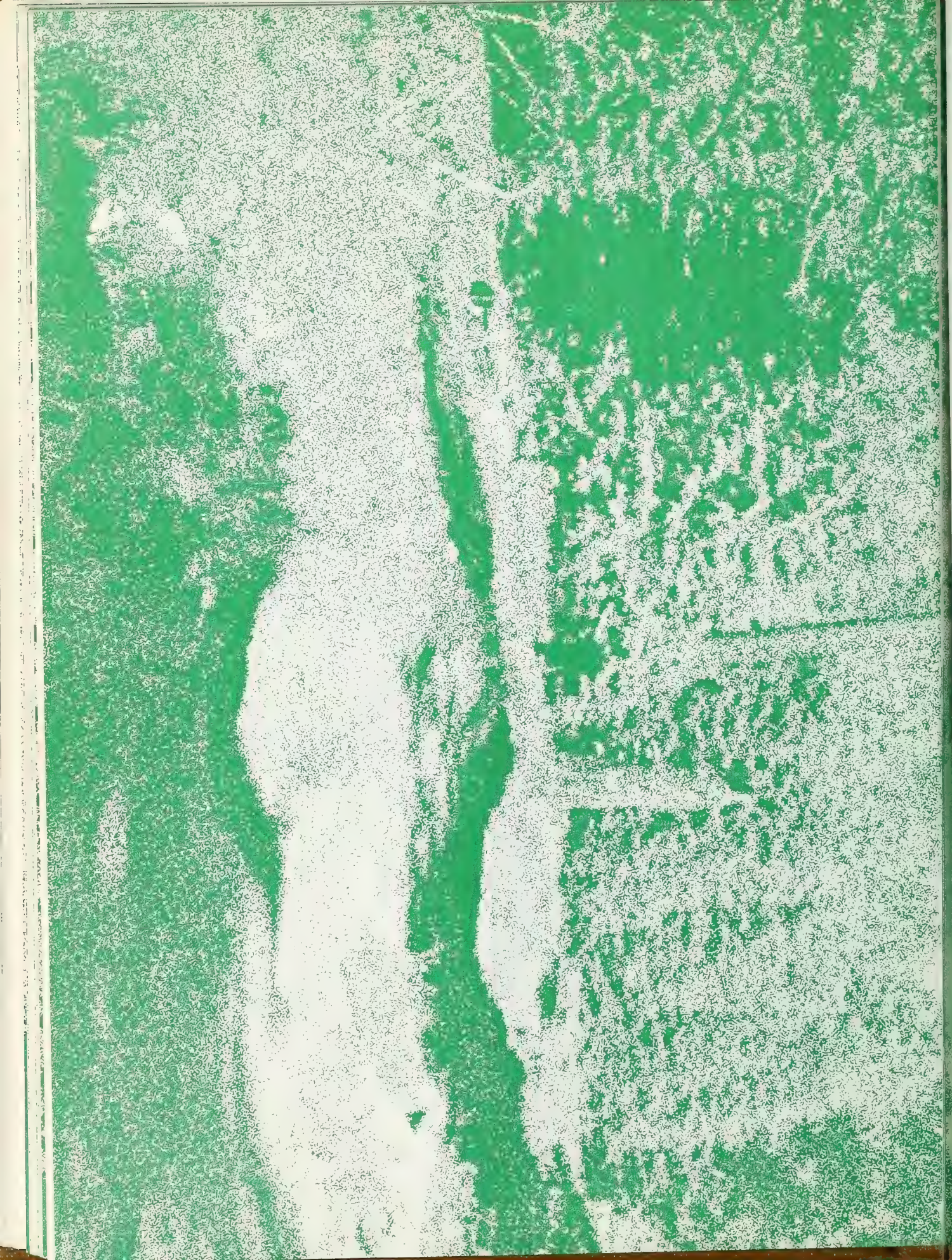
SOME EXTENSIONS OF THE SBE METHODOLOGY

This section presents some on-going esthetics-related research projects that offer possibilities for dealing more objectively with natural beauty as a forest resource. We invite critical comments and suggestions.
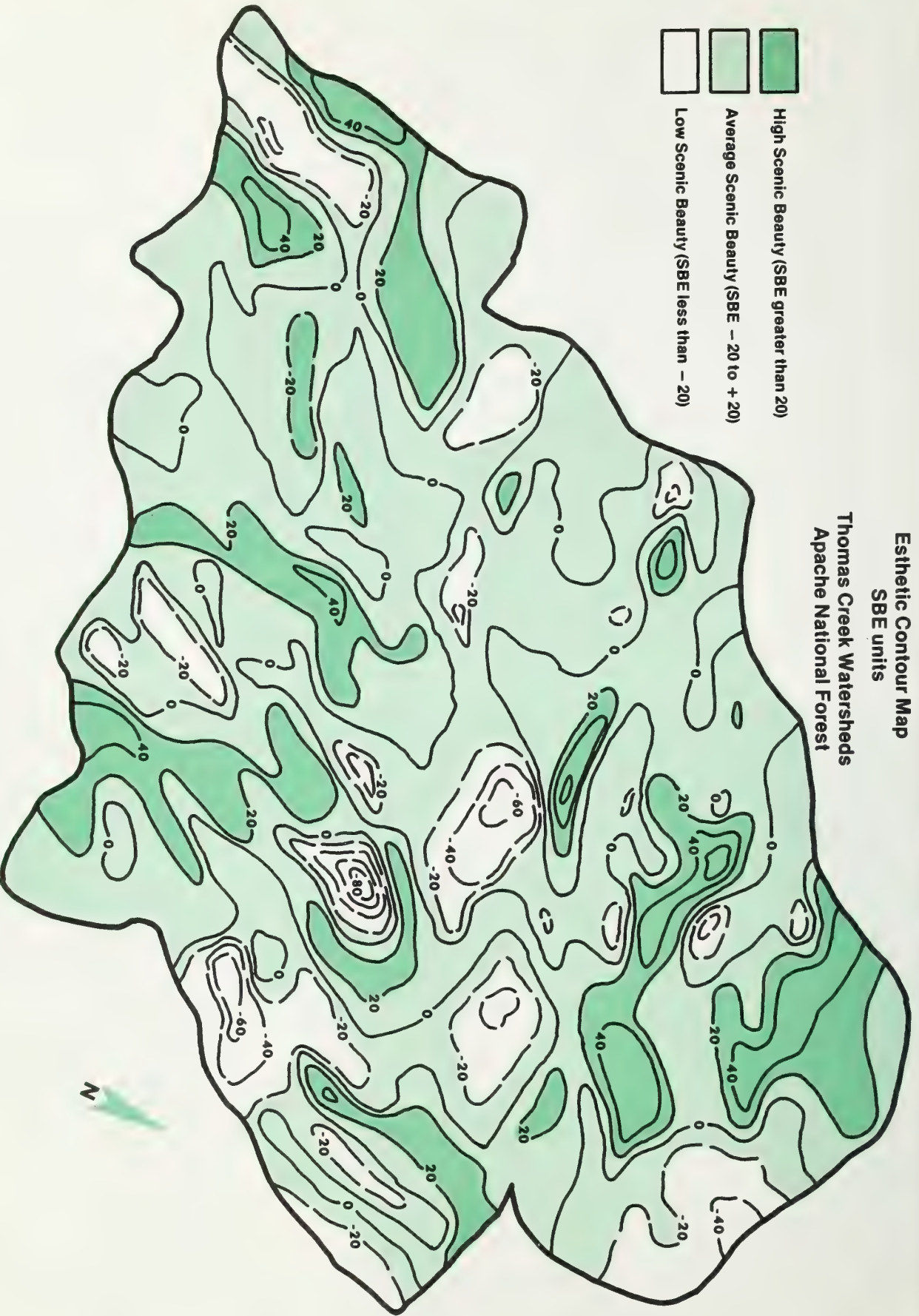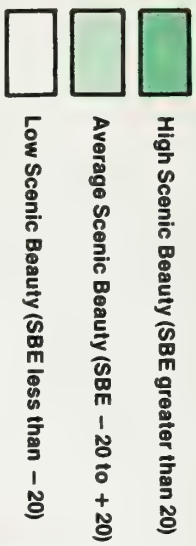
Esthetic Mapping

On page 54 is what might be called an esthetic or scenic beauty map. The area is the Thomas Creek Experimental Watersheds, which cover 1,073 acres on the Apache-Sitgreaves National Forest in Arizona. The contours represent lines of equal scenic quality; the zero (darker) contour represents the average scenic quality for these watersheds. The higher the number of a contour (darker areas), the more esthetically pleasing is the view from that line relative to the average; the lower the value (lighter areas) the less pleasing. Note that a contour on a topographic map represents the elevation at each point along that contour. In contrast, a contour on the scenic beauty map represents the scenic qualities of the view from each point on that contour.

This map was prepared in much the same way as a topographical map: SBEs were determined at specific ground points, and transferred to a map of the area. Contours were then drawn connecting points of equal value. To obtain the SBEs, randomly directed photos were taken at each of 250 equally-spaced stakes. The resulting slides were randomized and shown to groups of observers who rated them on scenic beauty. The SBE mathematical methodology had to be modified somewhat to permit this application; SBEs were computed for each stake based on the four slides.

We believe the concept of an esthetic contour map is valid, but we are concerned with and presently working on several important questions. For example, how close to one another do the sampling ground points need to be, and how many photos are necessary to adequately represent the scenic quality from a point? Answers to these questions for Thomas Creek will not necessarily hold for other areas. We are also uncertain of the obvious influences of dramatic changes in viewing

Esthetic Contour Map
SBE units
Thomas Creek Watersheds
Apache National Forest

High Scenic Beauty (SBE greater than 20)

Average Scenic Beauty (SBE −20 to +20)

Low Scenic Beauty (SBE less than −20)

depth across an area. None of these problems, however, seem insurmountable.

The design of scenic road and trail systems is but one application of such a map; the location of scenic vistas is another. The feasibility of overlaying a scenic beauty map with other resource maps offers the possibility of a direct interface with other resources and products. The effects of seasonal changes on esthetics could also be readily determined.
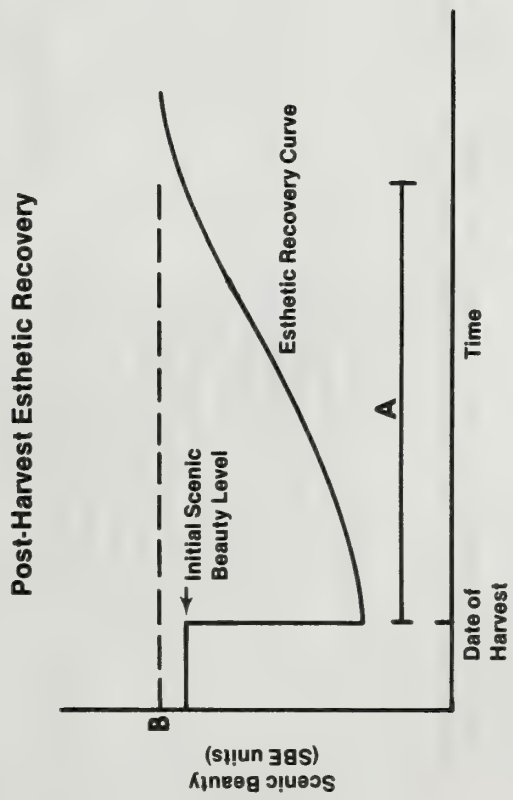
Time Sequence Studies

With few exceptions, scenic beauty is generally significantly poorer immediately following timber harvest than before. Over time, however, a site tends to recover toward (and may even exceed) preharvest esthetic levels. Postharvest recovery of scenic beauty may be illustrated by a simple time-dependent recovery curve (adjacent figure).

The two most relevant questions in this regard are: how much time ("A" in the figure) is required for scenic beauty to recover, and to what level ("B") does the area recover? The SBE Method may provide answers to these questions. We have been recording yearly recovery on several areas by slide sampling. All but one of the areas were cut just before sampling began in 1971; one area was cut just after our first sample in these areas. Control areas were established at the onset.

Our studies are designed to determine if the SBE Method is useful in answering the above two questions. We recognize that different areas and different species may recover at markedly different rates and that the nature of the harvest prescription is also an important influence (Shafer 1969). Nevertheless, the questions posed are important to forest decision-makers because they must justify decisions that will have long-range esthetic impacts.

Feature Analysis

Our major research effort is presently in the area of "Feature Analysis": the prediction of scenic quality from scaled landscape features. We have concentrated on **manageable** features or characteristics such

Post-Harvest Esthetic Recovery

Scenic Beauty (SBE units)

Initial Scenic Beauty Level

Esthetic Recovery Curve

A

Date of Harvest

Time

Selected Landscape Features and Their Relationship to Scenic Beauty and to Each Other

AMOUNT OF DOWNED WOOD – .87

- .89 Distribution of Downed Wood
- .87 Evidence of Manipulation of Downed Wood
- .84 Variability of Downed Wood
- .80 Density of Trees
- – .74 Tree Cover (Crown canopy)
- – .71 Proportion of Ponderosa Pine
- .71 Average Size of Downed Wood
- .67 Amount of Sky in Frame
- .63 Stumps
- – .62 Average Diameter of Trees

TREE DENSITY .74

- .95 Tree Cover
- – .94 Evidence of Manipulation of Downed Wood
- .92 Distribution of Downed Wood
- – .91 Variability of Downed Wood
- .89 Proportion of Ponderosa Pine
- .87 Dominant Visual Depth
- .82 Amount of Sky in Frame
- – .80 Amount of Rocks
- – .80 Amount of Downed Wood
- .76 Stumps
- .71 Amount of Shadow
- – .65 Average Size of Downed Wood
- – .62 Brightness
- – .60 Average Diameter of Trees

AVERAGE TREE DIAMETER .73

- .93 Modal Diameter of Trees
- .82 Variability of Density
- – .78 Amount of Rocks
- .76 Variability of Diameter of Trees
- .74 Proportion of Ponderosa Pine
- – .68 Stumps
- – .64 Brightness
- – .62 Variability of Downed Wood
- – .62 Amount of Downed Wood
- .60 Density of Trees

DISTRIBUTION OF DOWNED WOOD – .78

- .98 Evidence of Manipulation of Downed Wood
- – .92 Density of Trees
- .90 Amount of Downed Wood
- – .87 Variability of Downed Wood
- – .86 Tree Cover
- – .78 Proportion of Ponderosa Pine
- .76 Amount of Sky in Frame
- .74 Dominant Visual Depth
- .71 Average Size of Downed Wood
- .67 Amount of Rocks
- .65 General Slope of Terrain
- .64 Stumps

STUMPS – .60

- – .76 Density of Trees
- – .74 Proportion of Ponderosa Pine
- .72 Variability of Downed Wood
- – .72 Tree Cover (Crown canopy)
- .72 Amount of Rocks
- .70 Dominant Visual Depth
- .70 Evidence of Manipulation of Downed Wood
- – .68 Average Diameter of Trees
- .64 Distribution of Downed Wood
- .63 Amount of Downed Wood

CROWN-COVER CANOPY .61

- .95 Density of Trees
- – .87 Evidence of Manipulation of Downed Wood
- .87 Variability of Downed Wood
- – .86 Distribution of Downed Wood
- – .85 Amount of Sky in Frame
- – .84 Dominant Visual Depth
- .84 Proportion of Ponderosa Pine
- – .74 Amount of Downed Wood
- .74 Amount of Rocks
- – .72 Stumps
- – .69 Amount of Rocks
- – .68 Brightness
- – .65 Average Size of Downed Wood

Note: The numbers immediately to the right of the main (heading) variables are correlation coefficients between those independent variables and scenic beauty (using SBEs as the dependent variable). Intercorrelations between each subordinate variable and the main variable are indicated to the left. These data are preliminary and may only be applicable to Arizona ponderosa pine forests.

as tree density and downed wood. Several researchers — most notably Shafer and his colleagues — have related scenic beauty to landscape features. They have, however, often focused on variables over which management has little control

We initially drew up a list of manageable landscape features, hoping that a small number of these could eventually be combined into simple

equations (models) to accurately predict scenic beauty. The dependent variable of such models would, of course, be scenic beauty as estimated by SBEs. Because we had thousands of independently derived SBEs for 90 slides (those shown to the public groups discussed previously) we had the dependent variable accurately identified in a statistical sense. Scaled landscape features would be the "predictor" (independent variables). We employed some simple empirical (regression, correlation) techniques as well as factor and cluster analysis procedures in an attempt to (1) reduce the list of factors, and (2) determine those factors which are the best predictors of scenic beauty.

The adjacent table lists some of the factors we have been studying and some information as to how each correlates with scenic beauty and with the other variables. The evidence is clear that predictive models are feasible. Conceptually, such predictive models are not unlike other resource response models (hydrologic response models for instance) used to predict consequences (responses) of management options.

Once the models are derived, it would be unnecessary to test groups with slides except to update the models periodically to account for changes in public tastes and preferences; scenic quality changes would be estimated directly. Thus, for an array of management options, the decision-maker could compute estimates of public esthetic preferences for each alternative **before** a decision is made. This would reduce costs by eliminating the necessity for either on-site visitations or slide testing.

Two less obvious, but important, advantages of esthetic response models should be noted. The first is that, because the predicted SBEs are real numbers (interval outputs), they are easily integrated into comprehensive linear programming planning models. The other advantage is essentially economic: there is considerable overlap in the effort required to obtain the input data for the several resource response models. The opportunity to economize on data collection has long been recognized as an essential ingredient of comprehensive planning and management.
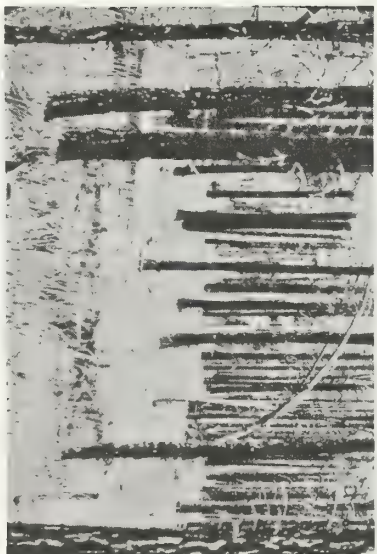
Amount of Downed Wood

Average Tree Density

Graphic Manipulation

Forests are extremely complex entities. Adequate experimental control of the type commonly exercised in laboratories or control plots is difficult to achieve because of the presence of so many confounding variables. In an effort to achieve the experimental control necessary for more accurately determining the effects of individual landscape features on scenic beauty, we are investigating several graphic representation techniques. The main objective is to be able to represent specific changes in landscape features (such as various amounts of downed wood) while holding the effects of other features (such as tree density, arrangement of downed wood, etc.) constant.

The adjacent pictures are two 4-part sequential, graphic manipulations of two important landscape features: downed wood and tree density. These representations were created by drawing with an opaque ink on Kodalith negatives and positives depending on whether a feature was to be removed or added. The Kodalith prints were made from high contrast black and white photographs shot from color slides. The only variable that changes in the first sequence is downed wood; in the second sequence, only tree density is varied. Thus we have accomplished in the laboratory what we could not in the field: experimental control over important variables.

We are, of course, aware of a major validity question: How well do these creations represent actual landscapes? This concern and others are being systematically considered in various experiments.

Other Uses and Extensions

Another area of investigation concerns applying the methodology to more accurately determine the effectiveness of Information and Education programs. Our initial work in this area is promising (Simpson et al., in press). We are also investigating the integration of the SBE approach and sophisticated "seen area" models such as VIEWIT (Amidon and Elsner 1968). Finally, we are coordinating our efforts with others involved with resource response models to define areas of overlapping data requirements.

SUMMARY & CONCLUSIONS

The problem of measuring scenic beauty has only recently become a major concern. The proliferation of techniques has been particularly dramatic during the past five years, no doubt reflecting a felt need by public land managers for a more substantive way to evaluate the scenic resource. One general criticism is that many techniques have been developed without adequate consideration of scientific criteria traditionally associated with measurement systems. Objective measurement would place scenic beauty, as a resource, on a more equal footing with other more tangible resources, and would also provide better justification for land use decisions. Also, the client-designer relationship between the public and landscape designers could be restored and improved.

The SBE Method measures public "perceptual preference" for various landscapes through a carefully defined system of rating color slides representative of these landscapes. It is based on the contention that scenic beauty judgments result from the interaction of observer perception and observer standards.

The SBE Method has evolved through numerous tests and applications. During development, particular attention was paid to commonly accepted criteria applicable to any measurement system. Results of application tests indicate that the method meets the criteria very well. The SBE index is a reliable and valid measure of perceived scenic beauty, and is applicable to a wide range of forest management problems.

The utility of the SBE Method is illustrated by the analysis of the reactions of user, interest, and professional groups. Differences and similarities in esthetic preference among the groups were apparent. Importantly, the relative strengths of their esthetic preferences were also revealed by the SBE measure. Further, differences in SBEs are indicative of true differences in the perceived scenic beauty of the landscapes, independent of differences in observers' standards.

Extensions and modifications of the SBE Method offer exciting possibilities for multiple-use forest management. Included in these possibilities are esthetic contour maps and prediction of scenic beauty based upon manageable landscape features. Although these applications are still in experimental and developmental stages, results so far are encouraging.

LITERATURE CITED

Amidon, Elliot, and Gary H. Elsner.
1968. Delineating landscape view areas. USDA For. Ser. Res. Note PSW-180, 5 p. Pacific Southwest For. and Range Exp. Stn., Berkeley, Cal.

Angus, R. C., and T. C. Daniel.
1974. Applying theory of signal detection in marketing product development and evaluation. Amer. J. Agric. Econ. 56(3):573-577.

Arthur, Louise M., and Ron S. Boster.
1976. Measuring scenic beauty: A selected annotated bibliography. USDA For. Serv. Gen. Tech. Rep. RM-25, Rocky Mt. For. and Range Exp. Stn., Fort Collins, Colo.

Bock, R. Darrell, and Lyle V. Jones.
1968. The measurement and prediction of judgement and choice. 370 p. Holden-Day, San Francisco.

Boster, Ron S.
[in press] Methodologies for scenic assessment. In: Proc. of a Conference/Workshop: Visual Quality and the Coastal Zone, Harper, David and John Warbach (Eds), College of Environmental Science and Forestry, State University of New York, Syracuse.

Boster, Ron S., and Terry C. Daniel.
1972. Measuring public responses to vegetative management. In: Proc. 16 Ann. Arizona Watershed Symp., Arizona Water Commission, Phoenix, p. 38-43.

Coughlin, Robert E., and Karen Goldstein.
1970. The extent of agreement among observers on environmental attractiveness. Reg. Sci. Res. Inst. Discussion Pap. 37, 37 p. Regional Sci. Res. Inst., Philadelphia, Pa.

Daniel, Terry C., Lawrence Wheeler, Ron Boster, and Paul Best.
1973. Quantitative evaluation of landscapes: an application of signal detection analysis to forest management alternatives. Man–Environment Systems 3(5):330-344.

Egan, J. P.
1958. Recognition memory and the operating characteristic. Technical Note AFCRC-TN-58-51, 31 p. Hearing Communication Laboratory, Indiana University.

Fabos, J. G.
1971. An analysis of environmental quality ranking systems in recreation. *In:* Recreation Symp. Proc., p. 40-55. Northeastern For. Exp. Stn., USDA For. Serv., Upper Darby, Pa.

Green, David M., and John A. Swets.
1966. Signal detection theory and psychophysics, 455 p. John Wiley, New York.

Hake, H. W., and A. S. Rodwan.
1966. Perception and recognition. *In:* Experimental Methods and Instrumentation in Psychology. p. 331-381, J. Sidowski (Ed), McGraw-Hill Book Co., New York.

Hays, William L.
1963. Statistics. 719 p. Holt, Rinehart and Winston, New York.

Hays, William L.
1969. Quantification in Psychology. 87 p. Brooks/Cole Publishing Co., Delmont, Cal.

Lee, W.
1969. Relationships between Thurstone category scaling and signal detection theory. Psychol. Bull. 71:101-107.

Leopold, Luna B.
1969. Quantitative comparison of some aesthetic factors among rivers. Geol. Survey Circ. 630:1-16. USDI Geol. Surv., Wash., D.C.

Litton, R. Burton, Jr.
1968. Forest landscape description and inventories — a basis for land planning and design. USDA For. Serv. Res. Pap. PSW-49. 64 p. Pacific Southwest For. and Range Exp. Stn., Berkeley, Cal.

Litton, R. Burton, Jr.
1974. Visual vulnerability of forest landscapes. J. For. 72(7):392-397.

Pearson, Henry A., and Donald A. Jameson.
1967. Relationship between timber and cattle production on ponderosa pine range: The Wild Bill Range. U.S. Dep. Agric., For. Serv., Rocky Mt. For. and Range Exp. Stn., Fort Collins, Colo., unnumbered publ., 10 p.

Potter, Dale R., and J. Alan Wagar.
1971. Techniques for inventorying manmade impact in roadway environments. USDA For. Serv. Res. Pap. PNW-121, 12 p. Pacific Northwest For. and Range Exp. Stn., Portland, Ore.

Redding, Martin J.
1973. Aesthetics in environmental planning. 187 p. Washington D.C.: U.S. Environmental Protection Agency. EPA-600/5-73-009.

Shafer, Elwood L., Jr
1969. Perception of natural environments. Environment and Behavior, 1(1):71–82.

Shafer, Elwood L., Jr., John F. Hamilton, Jr., and Elizabeth A. Schmidt.
1969. Natural landscape preferences: a predictive model. J. Leisure Res. 1(1):1–19.

Simpson, Cynthia J., Ted L. Rosenthal, Terry C. Daniel, and Glenn M. White.
[in press] Social influence variations in evaluating managed and unmanaged forest areas. J. Applied Psychol.

Swetts, J. A.
1973. The relative operating characteristic in psychology. Science 182(4116):990–1000.

Thurstone, L. L.
1927. A law of comparative judgement. Psychol. Review 34:278–286.

Thurstone, L. L.
1948. Psychophysical methods. In: Methods of Psychology. p. 124–157. T. G. Andrews (Ed), John Wiley, New York.

Torgerson, W. S.
1958. Theory and methods of scaling. 460 p. John Wiley, New York.

Turner, James M., and Frederic R. Larson.
1974. Cost analysis of experimental treatments on ponderosa pine watersheds. USDA For. Serv. Res. Pap. RM-116, 12 p. Rocky Mt. For. and Range Exp. Stn., Fort Collins, Colo.

U.S. Department of Agriculture.
1973. National Forest Landscape Management, Vol. 1, U.S. Dep. Agric., Agric. Handbook 434, 77 p.

U.S. Department of Agriculture.
1974. National Forest Landscape Management, Vol. 2, ch. 1: the Visual Management System. U.S. Dep. Agric., Agric. Handbook 462, 47 p.

Wheeler, L. T., T. C. Daniel, G. W. Seeley, and W. Swindell.
1971. Detectability of degraded visual signals: a basis for evaluating image-retrieval programs. Optical Sciences Center Tech. Rep. 73, Univ. of Ariz. 88 p. Tucson.

White, Glenn M.
1975. Contextual determinants of opinion judgments: field experimental probes of judgmental relativity boundary conditions. J. Personality and Social Psychol. 32:1047-1054.

Zube, Ervin H.
1974. Cross-disciplinary and intermode agreement on the description and evaluation of landscape resources. Environment and Behavior 6(1):69-89.

Zube, Ervin H., D. G. Pitt, and T. W. Anderson.
1974. Perception and measurement of scenic resources in the southern Connecticut River Valley. Institute for Man and His Environment, Publ. R-74-1. 191 p. Amherst, Mass.

Daniel, Terry C., and Ron S. Boster.
1976. Measuring landscape esthetics: The scenic beauty estima-
tion method. USDA For. Serv. Res. Pap. RM-167, 66 p. Rocky
Mt. For. and Range Exp. Stn., Fort Collins, Colo. 80521

The Scenic Beauty Estimation Method (SBE) provides quanti-
tative measures of esthetic preferences for alternative wildland
management systems. Extensive experimentation and testing with
user, interest, and professional groups validated the method. SBE
shows promise as an efficient and objective means for assessing
the scenic beauty of public forests and wildlands and also for pre-
dicting the esthetic consequences of alternative land uses. Exten-
sions and modifications of the basic methodology offer potentially
useful design, planning, and management tools.

Keywords: Esthetics, land use planning.

Daniel, Terry C., and Ron S. Boster.
1976. Measuring landscape esthetics: The scenic beauty estima-
tion method. USDA For. Serv. Res. Pap. RM-167, 66 p. Rocky
Mt. For. and Range Exp. Stn., Fort Collins, Colo. 80521

The Scenic Beauty Estimation Method (SBE) provides quanti-
tative measures of esthetic preferences for alternative wildland
management systems. Extensive experimentation and testing with
user, interest, and professional groups validated the method. SBE
shows promise as an efficient and objective means for assessing
the scenic beauty of public forests and wildlands and also for pre-
dicting the esthetic consequences of alternative land uses. Exten-
sions and modifications of the basic methodology offer potentially
useful design, planning, and management tools.

Keywords: Esthetics, land use planning.

Daniel, Terry C., and Ron S. Boster.
1976. Measuring landscape esthetics: The scenic beauty estima-
tion method. USDA For. Serv. Res. Pap. RM-167, 66 p. Rocky
Mt. For. and Range Exp. Stn., Fort Collins, Colo. 80521

The Scenic Beauty Estimation Method (SBE) provides quanti-
tative measures of esthetic preferences for alternative wildland
management systems. Extensive experimentation and testing with
user, interest, and professional groups validated the method. SBE
shows promise as an efficient and objective means for assessing
the scenic beauty of public forests and wildlands and also for pre-
dicting the esthetic consequences of alternative land uses. Exten-
sions and modifications of the basic methodology offer potentially
useful design, planning, and management tools.

Keywords: Esthetics, land use planning.

Daniel, Terry C., and Ron S. Boster.
1976. Measuring landscape esthetics: The scenic beauty estima-
tion method. USDA For. Serv. Res. Pap. RM-167, 66 p. Rocky
Mt. For. and Range Exp. Stn., Fort Collins, Colo. 80521

The Scenic Beauty Estimation Method (SBE) provides quanti-
tative measures of esthetic preferences for alternative wildland
management systems. Extensive experimentation and testing with
user, interest, and professional groups validated the method. SBE
shows promise as an efficient and objective means for assessing
the scenic beauty of public forests and wildlands and also for pre-
dicting the esthetic consequences of alternative land uses. Exten-
sions and modifications of the basic methodology offer potentially
useful design, planning, and management tools.

Keywords: Esthetics, land use planning.